



Salary Prediction (Regression)

The goal of this study is to train a model in order to predict a person's salary . The dataset used in this case study is found in <https://www.kaggle.com/datasets/mrsimple07/salary-prediction-data/data> and has 7 features and 1000 samples. It contains simulated data reflecting various factors influencing salary levels such as education, experience, location, job title, age, and gender.

The dataset contains no missing values and includes several categorical features. Categorical features contain multiple levels and the data was transformed to corresponding numeric codes, as detailed below:

Gender:

- Female (0)
- Male (1)

Education:

- High School (0)
- Bachelor (1)
- Master (2)
- PhD (3)

Location:

- Rural (0)
- Suburban (1)
- Urban (2)

Job_Title:

- Analyst (0)
- Director (1)
- Engineer (2)
- Manager (3)

Step 1: Import data from file

Right click on the input spreadsheet and choose the option "Import from file". Then navigate through your files to load the one with the salary data.

User Header	Col1	Col2	Col3	Col4	Col5	Col6
1	User Row ID					
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (D)	Col9	Col10	Col11
1	User Row ID	Education	Experience	Location	Job Title	Age	Gender	Salary			
2		0	8	2	3	63	1	84620.05366			
3		2	11	1	1	59	1	142591.2559			
4		1	28	1	3	61	0	97800.2554			
5		0	29	0	1	45	1	96834.67128			
6		3	25	2	0	26	0	132157.7862			
7		3	19	0	1	27	0	156312.9362			
8		3	4	0	1	60	0	130567.6495			
9		3	13	1	1	49	0	148707.7408			
10		1	20	2	2	25	0	95945.27543			
11		3	14	2	0	58	0	133339.3897			
12		1	23	0	3	23	0	101164.0924			
13		2	1	2	1	27	0	124251.9871			
14		0	11	2	3	63	1	35066.11435			
15		3	4	0	3	43	1	124972.5088			
16		2	7	1	1	44	0	126139.1367			
17		0	28	1	2	49	1	81007.19945			
18		0	4	1	0	37	1	60693.92471			
19		0	26	0	1	63	1	103386.0703			
20		2	7	2	1	53	0	140042.1389			
21		1	20	0	0	34	1	86048.11608			
22		2	3	0	2	62	0	73861.99295			

Step 2: Manipulate data

In order to use the data for training we have to exclude any columns that do not contain features. In our dataset there are no such columns. Therefore, we will include all columns in the training. We follow these steps to execute this:

- On the menu click on "Data Transformation" → "Data Manipulation" → "Select Column(s)"
- Select all columns.

The screenshot shows the 'Data Transformation' menu with options: Normalizers, Data Manipulation, Split, Variable Selection, Matrix Transpose, Sort by Column, and Fill Missing Column(s) Values. The 'Split' option is selected. Below the menu is a data table with 9 columns (Col1 to Col9) and 21 rows (User Header to User Row ID 21). The 'Select Column(s)' dialog box is open, showing 'Excluded Columns' (empty) and 'Included Columns' (Col2 -- Education, Col3 -- Experience, Col4 -- Location, Col5 -- Job_Title, Col6 -- Age, Col7 -- Gender, Col8 -- Salary). The 'Execute' button is highlighted.

The data will appear in the output spreadsheet.

Step 3: Split data

Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN_TEST_SPLIT" which we will use for splitting to create the train and test set.

Import data into the input spreadsheet of the "TRAIN_TEST_SPLIT" tab from the output of the "IMPORT" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

The screenshot shows the 'TRAIN_TEST_SPLIT' tab selected. The 'Import from SpreadSheet' option is highlighted. Below the menu is a data table with 9 columns (Col1 to Col9) and 21 rows (User Header to User Row ID 21). The 'Import from SpreadSheet' option is highlighted.

Split the dataset by choosing: "Data Transformation" → "Split" → "Random Partitioning". Then choose the "Training set percentage" and the column for the sampling as shown below:

Random Partitioning

Training set percentage: 75

☐ Usage of random generator seed: 612642324635300

☒ Stratified sampling: Col4 -- Location

Execute Cancel

User Header	Col1 (I)	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)
User Row ID	Education	Experience	Location	Job Title	Age	
1	0	8	2	3	63	
2	3	11	1	1	59	
3	1	28	1	3	61	
4	0	29	0	1	45	
5	3	25	2	0	26	
6	3	19	0	1	27	
7	3	4	0	1	60	
8	3	13	1	1	49	
9	1	20	2	2	25	
10	3	14	2	0	58	
11	1	23	0	3	23	
12	2	1	2	1	27	
13	0	11	2	3	63	
14	3	4	0	3	43	
15	2	7	1	1	44	
16	0	28	1	2	49	
17	0	4	1	0	37	
18	0	26	0	1	63	
19	2	7	2	1	53	
20	1	20	0	0	34	
21	2	3	0	2	62	

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8
User Row ID	Education	Experience	Location	Job Title	Age	Gender	Salary	
1	0	8	2	3	63	1	8462	
2	3	11	1	1	59	1	1425	
3	1	28	1	3	61	0	9780	
4	0	29	0	1	45	1	9683	
5	3	25	2	0	26	0	1321	
6	3	19	0	1	27	0	1563	
7	3	4	0	1	60	0	1305	
8	3	13	1	1	49	0	1487	
9	1	20	2	2	25	0	9594	
10	3	14	2	0	58	0	1333	
11	1	23	0	3	23	0	1011	
12	2	1	2	1	27	0	1242	
13	0	11	2	3	63	1	5506	
14	3	4	0	3	43	1	1249	
15	2	7	1	1	44	0	1261	
16	0	28	1	2	49	1	8100	
17	0	4	1	0	37	1	6069	
18	0	26	0	1	63	1	1033	
19	2	7	2	1	53	0	1400	
20	1	20	0	0	34	1	8604	
21	2	3	0	2	62	0	7386	

Step 4: Normalize the training set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALIZE_TRAIN_SET".

Import data into the input spreadsheet of the "NORMALIZE_TRAIN_SET" tab the train set from the output of the "TRAIN_TEST_SPLIT" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT : Training Set"

User Header	Col1	Col2 (Education)	Col3 (Experience)	Col4 (Location)	Col5 (Job_Title)	Col6 (Age)	Col7 (Gender)	Col8 (Salary)	Col9
1	0	8	2	3	63	1	0	94620.05366	
2	1	28	1	3	61	0	0	97800.2554	
3	3	25	2	0	26	0	0	132157.7862	
4	3	19	0	1	27	0	0	156312.9362	
5	3	13	1	1	49	0	0	148707.7408	
6	1	20	2	2	25	0	0	95945.27543	
7	3	14	2	0	58	0	0	133339.3897	
8	1	23	0	3	23	0	0	101164.0924	
9	2	1	2	1	27	0	0	124251.9871	
10	0	11	2	3	63	1	0	55066.11435	
11	3	4	0	3	43	1	0	124972.5088	
12	2	7	1	1	44	0	0	126139.1367	
13	0	28	1	2	49	1	0	81007.19945	
14	0	4	1	0	37	1	0	60693.92471	
15	0	26	0	1	63	1	0	103386.0703	
16	2	7	2	1	53	0	0	140042.1389	
17	1	20	0	0	34	1	0	86048.11608	
18	2	3	0	2	62	0	0	73861.99295	
19	3	19	1	0	62	1	0	106145.0055	
20	3	19	2	1	36	0	0	142981.5905	
21	2	7	1	0	21	0	0	96620.92939	

Normalize the data using Z-score by browsing: "Data Transformation" → "Normalizers" → "Z-Score". Then select all columns except "Salary" and click "Execute".

ZScore Normalizer

Excluded Columns: Col8 -- Salary

Included Columns: Col2 -- Education, Col3 -- Experience, Col4 -- Location, Col5 -- Job_Title, Col6 -- Age, Col7 -- Gender

Execute Cancel

The results will appear on the output spreadsheet.

File

Edit

Data Transformation

Analytics

Statistics

Plot

Help

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9
User Header	User Row ID	Education	Experience	Location	Job Title	Age	Gender	Salary	
1	0	8	2	3	63	1	84620.05366		
2	1	28	1	3	61	0	97800.2554		
3	3	25	2	0	26	0	132157.7862		
4	3	19	0	1	27	0	156312.9362		
5	3	13	1	1	49	0	148707.7408		
6	1	20	2	2	25	0	95945.27543		
7	3	14	2	0	58	0	133339.3897		
8	1	23	0	3	23	0	101164.0924		
9	2	1	2	1	27	0	124251.9871		
10	0	11	2	3	63	1	55066.11435		
11	3	4	0	3	43	1	124972.5088		
12	2	7	1	1	44	0	126139.1367		
13	0	28	1	2	49	1	81007.19945		
14	0	4	1	0	37	1	60693.92471		
15	0	26	0	1	63	1	103386.0703		
16	2	7	2	1	53	0	140042.1389		
17	1	20	0	0	34	1	86048.11608		
18	2	3	0	2	62	0	73861.99295		
19	3	19	1	0	62	1	106145.0055		
20	3	19	2	1	36	0	142981.5905		
21	2	7	1	0	21	0	96620.92939		

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
User Header	User Row ID	Education	Experience	Location	Job Title	Age	Gender	Salary
1	1	-1.306097775	-0.808659187	1.367874889	1.367874889	1.5315200494	0.9322895243	84620.05366
2	2	4581859	357434	673493	469699	0.9322895243	825693	97800.2554
3	3	-0.40407706	1.5762971261	0.0428123055	1.367874889	1.3829868425	-1.071197992	97800.2554
4	4	27285556	230034	11193825	673493	782078	198883	132157.7862
5	5	1.3891643520	1.2185512679	1.2777826567	-1.324248465	1.2180223277	-1.071197992	132157.7862
6	6	978048	00938	956325	0067823	6251279	198883	156312.9362
7	7	1.3891643520	0.5003667850	-1.192158045	-0.4353540680	-1.143711274	-1.071197992	156312.9362
8	8	978048	586069	773245	3487385	190747	198883	95945.27543
9	9	1.3891643520	-0.212420108	0.0428123055	-0.4353540680	0.4911528013	-1.071197992	95945.27543
10	10	978048	98732416	11193825	3487385	656355	198883	148707.7408
11	11	-0.40407706	0.6223146007	1.2777826567	0.4671671043	-1.292335281	-1.071197992	148707.7408
12	12	27285556	308208	956325	9939538	655699	198883	156312.9362
13	13	1.3891643520	-0.0931722293	1.2777826567	-1.324248465	1.1599608322	-1.071197992	156312.9362
14	14	978048	3133023	956325	0067823	750647	198883	132157.7862
15	15	-0.40407706	0.9800580477	-1.192158045	1.367874889	1.440959287	-1.071197992	132157.7862
16	16	27285556	52894	773245	673493	9282709	198883	95945.27543
17	17	0.4925436429	-1.643193897	1.2777826567	-0.4353540680	-1.143711274	-1.071197992	95945.27543
18	18	1247466	0755864	956325	3487385	190747	198883	81007.19945
19	19	-1.306097775	-0.450915740	1.2777826567	1.367874889	1.5315200494	0.9322895243	81007.19945
20	20	4581859	33356785	956325	673493	469699	825693	97800.2554
21	21	1.3891643520	-1.285650450	-1.192158045	1.367874889	0.0452807807	0.9322895243	97800.2554
22	22	978048	0933206	773245	673493	9349396	825693	132157.7862
23	23	0.4925436429	-0.927907003	0.0428123055	-0.4353540680	0.115992784	-1.071197992	132157.7862
24	24	1247466	014552	11193825	3487385	9373041	198883	156312.9362
25	25	-1.306097775	1.5762971261	0.0428123055	0.4671671043	0.4911528013	0.9322895243	156312.9362
26	26	4581859	230034	11193825	9939538	656355	825693	97800.2554
27	27	-1.306097775	-1.285650450	0.0428123055	-1.324248465	-0.400591239	0.9322895243	97800.2554
28	28	4581859	0535206	11193825	0067823	84693673	825693	132157.7862
29	29	-1.306097775	1.3378014947	-1.192158045	-0.4353540680	1.5315200494	0.9322895243	132157.7862

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

Step 5: Normalize the test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "NORMALIZE_TEST_SET".

Import data into the input spreadsheet of the "NORMALIZE_TEST_SET" tab the test set from the output of the "TRAIN_TEST_SPLIT" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet". From the available Select input tab options choose "TRAIN_TEST_SPLIT: Test Set".

File Edit Data Transformation Analytics Statistics Plot Help

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

NORMALIZE_TEST_SET

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9	Col10
1	3	11	1	1	59	1	142591.2559			
2	0	29	0	1	45	1	96834.67128			
3	3	4	0	1	60	0	130567.6495			
4	0	21	2	3	62	0	89426.94712			
5	1	12	2	3	20	1	104762.3985			
6	0	14	2	0	61	0	75559.24411			
7	3	24	1	1	23	0	153930.9339			
8	3	13	0	1	50	1	151123.3359			
9	2	7	0	1	40	1	106019.0204			
10	3	2	1	1	31	0	142122.6673			
11	0	19	2	3	62	1	95726.36719			
12	1	3	1	0	61	0	66216.70162			
13	3	13	1	2	35	0	127167.116			
14	3	2	0	0	20	0	96967.34243			
15	0	22	1	3	60	0	81946.59278			
16	3	22	2	1	56	1	169510.253			
17	2	17	0	3	62	0	136608.8081			
18	1	16	0	1	36	0	118098.3722			
19	2	17	0	0	34	1	109899.0494			
20	3	17	1	1	30	1	163022.8256			
21	2	11	0	0	40	0	98073.34046			

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

NORMALIZE_TEST_SET

User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							

Normalize the test set using the existing normalizer of the training set by browsing: "Analytics" → "Existing Model Utilization" → "Model (from Tab:) NORMALIZE_TRAIN_SET".

The screenshot shows the Isalos Analytics Platform interface. The 'Analytics' menu is open, showing options: Regression, Classification, Clustering, Anomaly Detection, and Existing Model Utilization. The 'Existing Model Execution' dialog is open, showing the 'Model' dropdown set to '(from Tab: NORMALIZE_TRAI...' and the 'Type' dropdown set to 'Z Score Normalizer Model'. The 'Description' field is empty. The 'Model Input' section lists the following mappings: Header -> Datatype, Education -> Double, Experience -> Double, Location -> Double, Job Title -> Double, Age -> Double, and Gender -> Double. The 'Transfer Column(s) to Output' checkbox is unchecked. The 'Execute' and 'Cancel' buttons are at the bottom.

User Header	Col1 (I)	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9
1	3	11	1	1	59	1	142591.2559		
2	0	29	0	1	45	1	96834.67128		
3	3	4	0	1	60	0	130567.6495		
4	0	21	2	3	62	0	89426.94712		
5	1	12	2	3	20	1	104762.3985		
6	0	14	2	0	61	0	75559.24411		
7	3	24	1	1	23	0	153930.9339		
8	3	13	0	1	50	1	151123.3359		
9	2	7	0	1	40	1	106019.0204		
10	3	2	1	1	31	0	142122.6673		
11	0	19	2	3	62	1	95726.36719		
12	1	3	1	0	61	0	66216.70162		
13	3	13	1	2	35	0	127167.116		
14	3	2	0	0	20	0	96967.34243		
15	0	22	1	3	60	0	81946.59278		
16	3	22	2	1	56	1	169510.253		
17	2	17	0	3	62	0	136608.8081		
18	1	16	0	1	36	0	118098.3722		
19	2	17	0	0	34	1	109899.0494		
20	3	17	1	1	30	1	163022.8256		
21	2	11	0	0	40	0	98073.34046		

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface. The 'Existing Model Execution' dialog is open, showing the 'Model' dropdown set to '(from Tab: NORMALIZE_TRAI...' and the 'Type' dropdown set to 'Z Score Normalizer Model'. The 'Description' field is empty. The 'Model Input' section lists the following mappings: Header -> Datatype, Education -> Double, Experience -> Double, Location -> Double, Job Title -> Double, Age -> Double, and Gender -> Double. The 'Transfer Column(s) to Output' checkbox is unchecked. The 'Execute' and 'Cancel' buttons are at the bottom.

User Header	Col1	Col2 (I)	Col3 (I)	Col4 (I)	Col5 (I)	Col6 (I)	Col7 (I)	Col8 (I)	Col9
1	3	11	1	1	59	1	142591.2559		
2	0	29	0	1	45	1	96834.67128		
3	3	4	0	1	60	0	130567.6495		
4	0	21	2	3	62	0	89426.94712		
5	1	12	2	3	20	1	104762.3985		
6	0	14	2	0	61	0	75559.24411		
7	3	24	1	1	23	0	153930.9339		
8	3	13	0	1	50	1	151123.3359		
9	2	7	0	1	40	1	106019.0204		
10	3	2	1	1	31	0	142122.6673		
11	0	19	2	3	62	1	95726.36719		
12	1	3	1	0	61	0	66216.70162		
13	3	13	1	2	35	0	127167.116		
14	3	2	0	0	20	0	96967.34243		
15	0	22	1	3	60	0	81946.59278		
16	3	22	2	1	56	1	169510.253		
17	2	17	0	3	62	0	136608.8081		
18	1	16	0	1	36	0	118098.3722		
19	2	17	0	0	34	1	109899.0494		
20	3	17	1	1	30	1	163022.8256		
21	2	11	0	0	40	0	98073.34046		

Step 6: Feature selection

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE_SELECTION_REGRESSION".

Import data into the input spreadsheet of the "FEATURE_SELECTION_REGRESSION" tab from the output of the "NORMALIZE_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

The screenshot shows the Isalos Analytics Platform interface. At the top, there's a menu bar with 'File', 'Edit', 'Data Transformation', 'Analytics', 'Statistics', 'Plot', and 'Help'. Below the menu, a workflow diagram shows the process: 'IMPORT' → 'TRAIN_TEST_SPLIT' → 'NORMALIZE_TRAIN_SET' → 'NORMALIZE_TEST_SET' → 'FEATURE_SELECTION_REGRESSION'. The main area displays a large data table with 15 rows and 9 columns. The columns are labeled 'User Header', 'Col1', 'Col2 (D)', 'Col3 (D)', 'Col4 (D)', 'Col5 (D)', 'Col6 (D)', 'Col7 (D)', and 'Col9'. The data includes numerical values for various features like Education, Experience, Location, Job Title, Age, Gender, and Salary.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col9
1		-1.300697775	-0.808659187	1.2777826567	1.3678748889	1.5315208494	0.9322895243	
2		4581859	3574334	956325	673493	469699	825693	
3		-0.404077066	1.5762971261	0.0428123055	1.3678748889	1.3828968425	-1.071197992	
4		27285556	230034	11193825	673493	782078	198883	
5		1.3891643520	1.2185536791	1.2777826567	-1.334248465	-1.218023277	-1.071197992	
6		978048	00938	956325	0067823	6251279	198883	
7		1.3891643520	0.5030667850	-1.192158045	-0.433540680	-1.143711274	-1.071197992	
8		978048	568069	773245	3487385	190747	198883	
9		1.3891643520	-0.212420108	0.0428123055	-0.433540680	0.4911528013	-1.071197992	
10		978048	98732416	11193825	3487385	656355	198883	
11		-0.404077066	0.6223146007	1.2777826567	0.4671671043	-1.292335281	-1.071197992	
12		27285556	308288	956325	0930536	059509	198883	
13		1.3891643520	-0.093172293	1.2777826567	-1.334248465	1.1599608322	-1.071197992	
14		978048	3133023	956325	0067823	750647	198883	
15		-0.404077066	0.9800580477	-1.192158045	1.3678748889	-1.440959287	-1.071197992	
16		27285556	528943	773245	673493	9282709	198883	
17		0.4925436429	-1.643393897	1.2777826567	-0.433540680	-1.143711274	-1.071197992	
18		1247466	0755864	956325	3487385	190747	198883	
19		1.300697775	-0.450915740	1.2777826567	1.3678748889	1.5315208494	0.9322895243	
20		4581859	33536785	956325	673493	469699	825693	
21		1.3891643520	-1.285650450	-1.192158045	1.3678748889	0.0452807807	0.9322895243	
22		978048	0535206	773245	673493	59349396	825693	
23		0.4925436429	-0.927907003	0.0428123055	-0.433540680	0.1195927841	-1.071197992	
24		1247466	0314552	11193825	3487385	9373041	198883	
25		-1.300697775	1.5762971261	0.0428123055	0.4671671043	0.4911528013	0.9322895243	
26		4581859	230034	11193825	0930536	656355	825693	
27		-1.300697775	-1.285650450	0.0428123055	-1.334248465	-0.400591239	0.9322895243	
28		4581859	0535206	11193825	0067823	84693673	825693	
29		-1.300697775	1.378014947	-1.192158045	-0.433540680	1.5315208494	0.9322895243	
30		4581859	7406	773245	1487385	860609	8374603	

Choose the most important features using the Regression Analysis by browsing: "Data Transformation" → "Variable Selection" → "Regression Analysis". Then choose the "Salary" column as the intercept column, the Significance level (α) as 0.05 and include all columns.

The screenshot shows the 'Regression Analysis Model' dialog box. It has a 'Significance Level (α)' field set to 0.06. The 'Select Intercept Column' dropdown is set to 'Col8 -- Salary'. The 'Excluded Columns' list is empty. The 'Included Columns' list contains: 'Col2 -- Education', 'Col3 -- Experience', 'Col4 -- Location', 'Col5 -- Job Title', 'Col6 -- Age', and 'Col7 -- Gender'. There are 'Execute' and 'Cancel' buttons at the bottom.

The results will appear on the output spreadsheet.

File Edit Data Transformation Analytics Statistics Plot Help									
	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
User Header	User Row ID	Education	Experience	Location	Job Title	Age	Gender	Salary	
1	-1.300697775	3574334	956325	1.2777826567	1.3678748889	1.5315200494	0.9322895243	84620.05366	
2	4581859	3574334	956325	1.2777826567	1.3678748889	1.5315200494	0.9322895243	84620.05366	
3	-0.404077066	1.5762971261	0.0428123055	1.3678748889	1.3828968425	-1.071197992	132157.7862		
4	27285556	230034	11193825	673493	782078	198883	156312.9362		
5	1.3891643520	1.2185536791	1.2777826567	-1.334248465	-1.218023277	-1.071197992	148707.7408		
6	978048	00938	956325	0067823	6251279	198883	95945.27543		
7	1.3891643520	-1.192158045	0.0336407850	-0.433540680	-1.143711274	-1.071197992	101164.0924		
8	978048	568069	773245	3487385	190747	198883	971856		
9	1.3891643520	-0.212420108	0.0428123055	-0.433540680	0.4911528013	-1.071197992	101164.0924		
10	978048	98732416	11193825	3487385	656355	198883	124251.9871		
11	-0.404077066	0.6223146007	1.2777826567	0.4671671043	-1.292335281	-1.071197992	101164.0924		
12	27285556	308288	956325	0930536	059509	198883	133239.3897		
13	1.3891643520	-0.093172293	1.2777826567	-1.334248465	1.1599608322	-1.071197992	133239.3897		
14	978048	3133023	956325	0067823	750647	198883	133239.3897		
15	-0.404077066	0.9800580477	-1.192158045	1.3678748889	-1.440959287	-1.071197992	101164.0924		
16	27285556	528943	773245	673493	9282709	198883	101164.0924		
17	0.4925436429	-1.643193897	1.2777826567	-0.433540680	-1.143711274	-1.071197992	124251.9871		
18	1247466	0755864	956325	3487385	190747	198883	124251.9871		
19	-1.300697775	-0.450915740	1.2777826567	1.3678748889	1.5315200494	0.9322895243	55066.11435		
20	4581859	33536785	956325	673493	469699	825693	124972.5088		
21	1.3891643520	-1.285650450	-1.192158045	1.3678748889	0.0452807807	0.9322895243	126139.1367		
22	978048	0535206	773245	673493	59349396	825693	124972.5088		
23	0.4925436429	-0.927907003	0.0428123055	-0.433540680	0.1195927841	-1.071197992	126139.1367		
24	1247466	0314552	11193825	3487385	9373041	198883	126139.1367		
25	-1.300697775	1.5762971261	0.0428123055	0.4671671043	0.4911528013	0.9322895243	126139.1367		
26	4581859	230034	11193825	0930536	656355	825693	126139.1367		
27	-1.300697775	-1.285650450	0.0428123055	-1.334248465	-0.400591239	0.9322895243	126139.1367		
28	4581859	0535206	11193825	0067823	84693673	825693	126139.1367		
29	-1.300697775	1.3778014947	-1.192158045	-0.433540680	1.5315200494	0.9322895243	126139.1367		
30	4581859	7496	773245	3487385	469699	825693	126139.1367		
31	0.4925436429	-0.927907003	1.2777826567	-0.433540680	0.7884008151	-1.071197992	126139.1367		
32	1247466	0314552	956325	3487385	031596	198883	126139.1367		
33	-0.404077066	0.6223146007	-1.192158045	-1.334248465	-0.623527250	0.9322895243	126139.1367		
34	27285556	308288	773245	0067823	1500798	825693	126139.1367		
35	0.4925436429	-1.404898265	-1.192158045	0.4671671043	1.4572088460	-1.071197992	126139.1367		
36	1247466	7275426	773245	0930536	125887	198883	126139.1367		
37	1.3891643520	0.5030667850	0.0428123055	-1.334248465	1.4572088460	0.9322895243	126139.1367		
38	978048	568069	11193825	0067823	125887	825693	126139.1367		
39	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
40	978048	568069	11193825	0067823	125887	825693	126139.1367		
41	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
42	978048	568069	11193825	0067823	125887	825693	126139.1367		
43	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
44	978048	568069	11193825	0067823	125887	825693	126139.1367		
45	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
46	978048	568069	11193825	0067823	125887	825693	126139.1367		
47	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
48	978048	568069	11193825	0067823	125887	825693	126139.1367		
49	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
50	978048	568069	11193825	0067823	125887	825693	126139.1367		
51	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
52	978048	568069	11193825	0067823	125887	825693	126139.1367		
53	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
54	978048	568069	11193825	0067823	125887	825693	126139.1367		
55	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
56	978048	568069	11193825	0067823	125887	825693	126139.1367		
57	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
58	978048	568069	11193825	0067823	125887	825693	126139.1367		
59	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
60	978048	568069	11193825	0067823	125887	825693	126139.1367		
61	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
62	978048	568069	11193825	0067823	125887	825693	126139.1367		
63	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
64	978048	568069	11193825	0067823	125887	825693	126139.1367		
65	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
66	978048	568069	11193825	0067823	125887	825693	126139.1367		
67	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
68	978048	568069	11193825	0067823	125887	825693	126139.1367		
69	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
70	978048	568069	11193825	0067823	125887	825693	126139.1367		
71	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
72	978048	568069	11193825	0067823	125887	825693	126139.1367		
73	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
74	978048	568069	11193825	0067823	125887	825693	126139.1367		
75	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
76	978048	568069	11193825	0067823	125887	825693	126139.1367		
77	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
78	978048	568069	11193825	0067823	125887	825693	126139.1367		
79	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
80	978048	568069	11193825	0067823	125887	825693	126139.1367		
81	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
82	978048	568069	11193825	0067823	125887	825693	126139.1367		
83	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
84	978048	568069	11193825	0067823	125887	825693	126139.1367		
85	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
86	978048	568069	11193825	0067823	125887	825693	126139.1367		
87	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
88	978048	568069	11193825	0067823	125887	825693	126139.1367		
89	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
90	978048	568069	11193825	0067823	125887	825693	126139.1367		
91	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
92	978048	568069	11193825	0067823	125887	825693	126139.1367		
93	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
94	978048	568069	11193825	0067823	125887	825693	126139.1367		
95	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
96	978048	568069	11193825	0067823	125887	825693	126139.1367		
97	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
98	978048	568069	11193825	0067823	125887	825693	126139.1367		
99	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
100	978048	568069	11193825	0067823	125887	825693	126139.1367		
101	1.3891643520	0.5030667850	1.2777826567	-0.433540680	-0.474903243	-1.071197992	126139.1367		
102	978048	568069	11193825	0067823	125887	825693	126139.1367		
103	1.3891643520	0.5030667850							

The significant features according to the p-value are the following:

- Salary (p-value = 0.0)
- Education (p-value = 1.1188205679825185E-203)
- Experience (p-value = 7.035817529201593E-54)
- Location (p-value = 1.7701240587180506E-13)
- Job_Title (p-value = 1.0075657611706519E-8)

Step 7: Feature selection: train set

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE_SELECTION_TRAIN_SET".

Import data into the input spreadsheet of the "FEATURE_SELECTION_TRAIN_SET" tab from the output of the "NORMALIZE_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

FileEditData TransformationAnalyticsStatisticsPlotHelp

IMPORT

TRAIN_TEST_SPLIT

NORMALIZE_TRAIN_SET

NORMALIZE_TEST_SET

FEATURE_SELECTION_REGRESSION

FEATURE_SELECTION_TRAIN_SET

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	-1.300697775	0.808659187	1.2777826567	1.3678748889	1.5315200494	0.9322895243	84620.05366	
2	4581859	3574334	956325	673493	469699	825693	97800.2554	
3	-0.404077066	1.5762971261	0.0428123055	1.3678748889	1.3828968425	-1.071197992	132157.7862	
4	27285556	230034	11193825	673493	782078	198883	156312.9362	
5	1.3891643520	1.2185536791	1.2777826567	-1.334248465	-1.218023277	-1.071197992	148707.7408	
6	978048	00938	956325	0067823	6251279	198883	95945.27543	
7	1.3891643520	0.5030667850	-1.192158045	-0.433540680	190747	-1.071197992	133339.3897	
8	978048	568069	773245	3487385	750647	198883	124251.9871	
9	1.3891643520	-0.212420108	0.0428123055	-0.433540680	0.4911528013	-1.071197992	101164.0924	
10	978048	98732416	11193825	3487385	656355	198883	124972.5088	
11	-0.404077066	0.6223146007	1.2777826567	0.4671671043	-1.292335281	-1.071197992	126139.1367	
12	27285556	308288	956325	0930536	059509	198883	126139.1367	
13	1.3891643520	-0.093172293	1.2777826567	-1.334248465	1.1599608322	-1.071197992	126139.1367	
14	978048	3133023	956325	0067823	750647	198883	126139.1367	
15	-0.404077066	0.9800580477	-1.192158045	1.3678748889	-1.440959287	-1.071197992	126139.1367	
16	27285556	528943	773245	673493	9282709	198883	126139.1367	
17	0.4925436429	-1.643193897	1.2777826567	-0.433540680	-1.143711274	-1.071197992	126139.1367	
18	1247466	0755864	956325	3487385	190747	198883	126139.1367	
19	-1.300697775	-0.450915740	1.2777826567	1.3678748889	1.5315200494	0.9322895243	126139.1367	
20	4581859	33536785	956325	673493	469699	825693	126139.1367	
21	1.3891643520	-1.285650450	-1.192158045	1.3678748889	0.0452807807	0.9322895243	126139.1367	
22	978048	0535206	773245	673493	59349396	825693	126139.1367	
23	0.4925436429	-0.927907003	0.0428123055	-0.433540680	0.1195927841	-1.071197992	126139.1367	
24	1247466	0314552	11193825	3487385	9373041	198883	126139.1367	
25	-1.300697775	1.5762971261	0.0428123055	0.4671671043	0.4911528013	0.9322895243	126139.1367	

User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								

IMPORTTRAIN_TEST_SPLITNORMALIZE_TRAIN_SETNORMALIZE_TEST_SETFEATURE_SELECTION_REGRESSIONFEATURE_SELECTION_TRAIN_SET

Manipulate the data by choosing the columns that correspond to the significant features (from the previous step): "Data Transformation" → "Data Manipulation" → "Select Column(s)".

The screenshot shows the 'Data Manipulation' menu with options: Remove Column(s), Select Column(s), Split, Variable Selection, Matrix Transpose, Sort by Column, and Fill Missing Column(s) Values. The 'Select Column(s)' dialog box is open, showing 'Excluded Columns' (Col6 -- Age, Col7 -- Gender) and 'Included Columns' (Col2 -- Education, Col3 -- Experience, Col4 -- Location, Col5 -- Job_Title, Col8 -- Salary). The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	-1.300697775	-0.808659187	1.2777826567	1.3678748889	1.5315200494	0.9322895243	84620.05366	
2	4581859	3574334	956325	673493	469699	825693	97800.2554	
3	1.3891643520	1.2185536791	1.2777826567	-1.334248465	-1.218023277	-1.071197992	132157.7862	
4	978048	00938	956325	0067823	6251279	198883	156312.9362	
5	1.3891643520	0.5030667850	-1.192158045	-0.433540680	0.4911528013	-1.071197992	148707.7408	
6	978048	98732416	11193825	3487385	656355	198883	95945.27543	
7	-0.404077066	0.6223146007	1.2777826567	0.4671671043	-1.292335281	-1.071197992	133339.3897	
8	1.3891643520	-0.093172293	1.2777826567	-1.334248465	1.1599608322	-1.071197992	124251.9871	
9	978048	3133023	956325	0067823	750647	198883	55066.11435	
10	-0.404077066	0.9800580477	-1.192158045	1.3678748889	-1.440959287	-1.071197992	101164.0924	
11	27285556	528943	773245	673493	9282709	198883	124972.5088	
12	0.4925436429	-1.643393897	1.2777826567	-0.433540680	-1.143711274	-1.071197992	126139.1367	
13	1247466	0755864	956325	3487385	190747	198883	81007.19945	

The results will appear on the output spreadsheet.

The screenshot shows the 'Feature Selection' workflow in the 'Data Transformation' tab. The 'Import from Spreadsheet' button is highlighted. The resulting output spreadsheet is shown, displaying the selected features (Education, Experience, Location, Job_Title, Age, Gender, Salary) for the test set.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)
1	-1.300697775	-0.808659187	1.2777826567	1.3678748889	1.5315200494	0.9322895243	84620.05366	
2	4581859	3574334	956325	673493	469699	825693	97800.2554	
3	1.3891643520	1.2185536791	1.2777826567	-1.334248465	-1.218023277	-1.071197992	132157.7862	
4	978048	00938	956325	0067823	6251279	198883	156312.9362	
5	1.3891643520	0.5030667850	-1.192158045	-0.433540680	0.4911528013	-1.071197992	148707.7408	
6	978048	98732416	11193825	3487385	656355	198883	95945.27543	
7	-0.404077066	0.6223146007	1.2777826567	0.4671671043	-1.292335281	-1.071197992	133339.3897	
8	1.3891643520	-0.093172293	1.2777826567	-1.334248465	1.1599608322	-1.071197992	124251.9871	
9	978048	3133023	956325	0067823	750647	198883	55066.11435	
10	-0.404077066	0.9800580477	-1.192158045	1.3678748889	-1.440959287	-1.071197992	101164.0924	
11	27285556	528943	773245	673493	9282709	198883	124972.5088	
12	0.4925436429	-1.643393897	1.2777826567	-0.433540680	-1.143711274	-1.071197992	126139.1367	
13	1247466	0755864	956325	3487385	190747	198883	81007.19945	

Step 8: Feature selection: test set

Create a new tab by pressing the "+" button on the bottom of the page with the name "FEATURE_SELECTION_TEST_SET".

Import data into the input spreadsheet of the "FEATURE_SELECTION_TEST_SET" tab from the output of the "NORMALIZE_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
1	1.3891643520	-0.450915740	0.0428123055	-0.433540680	1.2342728357	0.9322895243	142591.2559		
2	-1.300697775	1.6955449417	-1.192158045	-0.433540680	0.1939047876	0.9322895243	96834.67128		
3	1.3891643520	-1.285650450	-1.192158045	-0.433540680	1.3085848391	-1.071197992	130567.6495		
4	-1.300697775	0.7415624164	0.0428123055	-0.433540680	1.4572088460	-1.071197992	89426.94712		
5	-0.404077066	-0.331667924	1.2777826567	1.3678748889	-1.663895298	0.9322895243	104762.3985		
6	-1.300697775	-0.093172293	1.2777826567	-1.334248465	1.3828968425	-1.071197992	75559.24411		
7	1.3891643520	1.0993058634	0.0428123055	-0.433540680	-1.440959287	-1.071197992	153930.9339		
8	1.3891643520	-0.212420108	-1.192158045	-0.433540680	0.5654648048	0.9322895243	151123.3359		
9	0.4925436429	-0.927907003	-1.192158045	-0.433540680	-0.177655229	0.9322895243	106019.0204		
10	1.3891643520	-1.524146081	0.0428123055	-0.433540680	-0.846463260	-1.071197992	142122.6673		
11	-1.300697775	0.5030667850	1.2777826567	1.3678748889	1.4572088460	0.9322895243	95726.36719		
12	-0.404077066	-1.404898265	0.0428123055	-1.334248465	1.3828968425	-1.071197992	66216.70162		
13	1.3891643520	-0.212420108	0.0428123055	0.4671671043	-0.549215246	-1.071197992	127167.116		

Manipulate the data by choosing the columns that correspond to the significant features (from step 7): “Data Transformation” → “Data Manipulation” → “Select Column(s)”.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8
1	1.3891643520	-0.450915740	0.0428123055	-0.433540680	1.2342728357	0.9322895243	142591.	
2	-1.300697775	1.6955449417	-1.192158045	-0.433540680	0.1939047876	0.9322895243	96834.6	
3	1.3891643520	-1.285650450	-1.192158045	-0.433540680	1.3085848391	-1.071197992	130567.	
4	-1.300697775	0.7415624164	0.0428123055	-0.433540680	1.4572088460	-1.071197992	89426.9	
5	-0.404077066	-0.331667924	1.2777826567	1.3678748889	-1.663895298	0.9322895243	104762.	
6	-1.300697775	-0.093172293	1.2777826567	-1.334248465	1.3828968425	-1.071197992	75559.2	
7	1.3891643520	1.0993058634	0.0428123055	-0.433540680	-1.440959287	-1.071197992	153930.	
8	1.3891643520	-0.212420108	-1.192158045	-0.433540680	0.5654648048	0.9322895243	151123.	
9	0.4925436429	-0.927907003	-1.192158045	-0.433540680	-0.177655229	0.9322895243	106019.	
10	1.3891643520	-1.524146081	0.0428123055	-0.433540680	-0.846463260	-1.071197992	142122.	
11	-1.300697775	0.5030667850	1.2777826567	1.3678748889	1.4572088460	0.9322895243	95726.3	
12	-0.404077066	-1.404898265	0.0428123055	-1.334248465	1.3828968425	-1.071197992	66216.7	
13	1.3891643520	-0.212420108	0.0428123055	0.4671671043	-0.549215246	-1.071197992	127167.	

Excluded Columns

Col6 -- Age
Col7 -- Gender

Included Columns

Col2 -- Education
Col3 -- Experience
Col4 -- Location
Col5 -- Job Title
Col8 -- Salary

Execute Cancel

The results will appear on the output spreadsheet.

Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)	Col8 (D)	Col9
User Header	User Row ID	Education	Experience	Location	Job Title	Age	Gender	Salary
1		1.3891643520	-0.450915740	0.0428123055	-0.433540680	1.2342728357	0.9322895243	142591.2559
2		978048	33536785	11193825	3487385	094457	825693	
3		-1.300697775	1.6955449417	-1.192158045	-0.433540680	0.1939047876	0.9322895243	96834.67128
4		4581859	970254	773245	3487385	2811145	825693	130567.6495
5		1.3891643520	-1.285650450	-1.192158045	-0.433540680	1.3085848391	-1.071197992	
6		978048	0535206	773245	3487385	430268	198883	
7		-1.300697775	0.7415624164	1.2777826567	1.3678748889	-1.4572088460	-1.071197992	89426.94712
8		4581859	048505	956325	673493	125887	198883	
9		-0.404077066	-0.331667924	1.2777826567	1.3678748889	-1.663895298	0.9322895243	104762.3985
10		27285556	661346	956325	673493	2314141	825693	
11		-1.300697775	-0.093172293	1.2777826567	-1.334248465	1.382968425	-1.071197992	
12		4581859	3133023	956325	0067823	782078	198883	75559.24411
13		1.3891643520	1.0993058634	0.0428123055	-0.433540680	-1.440959287	-1.071197992	
14		978048	269162	11193825	3487385	9282709	198883	153930.9339
15		1.3891643520	-0.212420108	-1.192158045	-0.433540680	0.5654648048	0.9322895243	
16		978048	98732416	773245	3487385	000165	825693	151123.3359
17		0.4925436429	-0.927907003	-1.192158045	-0.433540680	-0.177555229	0.9322895243	
18		1247466	0314552	773245	3487385	54379365	825693	106019.0204
19		1.3891643520	-1.524146081	0.0428123055	-0.433540680	-0.846463260	-1.071197992	
20		978048	4015644	11193825	3487385	4532228	198883	142122.6673
21		-1.300697775	0.5030667850	1.2777826567	1.3678748889	1.4572088460	0.9322895243	
22		4581859	568069	956325	673493	125887	825693	95726.36719
23		-0.404077066	-1.404898265	0.0428123055	-1.334248465	1.382968425	-1.071197992	
24		27285556	7275426	11193825	0067823	782078	198883	66216.70162
25		1.3891643520	-0.212420108	0.0428123055	-0.433540680	-0.549215246	-1.071197992	127167.116

Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7
User Header	User Row ID	Education	Experience	Location	Job Title	Salary
1		1.3891643520	-0.450915740	0.0428123055	-0.433540680	142591.2559
2		978048	33536785	11193825	3487385	
3		-1.300697775	1.6955449417	-1.192158045	-0.433540680	96834.67128
4		4581859	970254	773245	3487385	130567.6495
5		1.3891643520	-1.285650450	-1.192158045	-0.433540680	
6		978048	0535206	773245	3487385	
7		-1.300697775	0.7415624164	1.2777826567	1.3678748889	89426.94712
8		4581859	048505	956325	673493	
9		-0.404077066	-0.331667924	1.2777826567	1.3678748889	104762.3985
10		27285556	661346	956325	673493	
11		-1.300697775	-0.093172293	1.2777826567	-1.334248465	75559.24411
12		1.3891643520	1.0993058634	0.0428123055	-0.433540680	
13		978048	269162	11193825	3487385	153930.9339
14		1.3891643520	-0.212420108	-1.192158045	-0.433540680	
15		978048	98732416	773245	3487385	151123.3359
16		0.4925436429	-0.927907003	-1.192158045	-0.433540680	
17		1247466	0314552	773245	3487385	106019.0204
18		1.3891643520	-1.524146081	0.0428123055	-0.433540680	
19		978048	4015644	11193825	3487385	142122.6673
20		-1.300697775	0.5030667850	1.2777826567	1.3678748889	
21		4581859	568069	956325	673493	95726.36719
22		-0.404077066	-1.404898265	0.0428123055	-1.334248465	
23		27285556	7275426	11193825	0067823	66216.70162
24		1.3891643520	-0.212420108	0.0428123055	-0.433540680	127167.116

Step 9: Train the model

Create a new tab by pressing the "+" button on the bottom of the page with the name "TRAIN_MODEL(.fit)".

Import data into the input spreadsheet of the "TRAIN_MODEL(.fit)" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
User Header	User Row ID	Education	Experience	Location	Job Title	Age	Gender	Salary
1		-1.300697775	-0.808659187	1.2777826567	1.3678748889	04620.05366		
2		4581859	3574334	956325	673493			
3		-0.404077066	1.5762971261	0.0428123055	1.3678748889	97800.2554		
4		27285556	230034	11193825	673493			
5		1.3891643520	1.2185536791	1.2777826567	-1.334248465	132157.7862		
6		978048	00938	956325	0067823			
7		1.3891643520	0.5030667850	-1.192158045	-0.433540680	156312.9362		
8		978048	568069	773245	3487385			
9		1.3891643520	-0.212420108	0.0428123055	-0.433540680	148707.7408		
10		978048	98732416	11193825	3487385			
11		-0.404077066	0.6223146007	1.2777826567	0.4671671043	95945.27543		
12		27285556	308288	956325	0930536			
13		1.3891643520	-0.093172293	1.2777826567	-1.334248465	133339.3897		
14		978048	3133023	956325	0067823			
15		-0.404077066	0.9800580477	-1.192158045	1.3678748889	101164.0924		
16		27285556	528943	773245	673493			
17		0.4925436429	-1.643393897	1.2777826567	-0.433540680	124251.9871		
18		1247466	0755864	956325	3487385			
19		-1.300697775	-0.450915740	0.0428123055	1.3678748889	55066.11435		
20		4581859	33536785	11193825	3487385			
21		1.3891643520	-1.285650450	-1.192158045	1.3678748889	124972.5088		
22		978048	0535206	773245	673493			
23		0.4925436429	-0.927907003	0.0428123055	-0.433540680	126139.1367		
24		1247466	0314552	11193825	3487385			
25		-1.300697775	1.5762971261	0.0428123055	0.4671671043	11077.19045		

Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
User Header	User Row ID	Education	Experience	Location	Job Title	Salary	
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							

Use the k Nearest Neighbors (kNN) method to train and fit the model by browsing: "Analytics" → "Regression" → "k Nearest Neighbors (kNN)" and set the "Target Column" as the column corresponding to "Salary" and the "Number of Neighbors" to 3.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7
1	1.300697775	-0.808659187	1.2777826567	1.3678748889	84620.05366		
2	4581859	3574334	956325	673493	97800.2554		
3	-0.404077066	1.5762971261	0.0428123055	1.3678748889	132157.7862		
4	1.3891643520	1.2185536791	1.1193825	-0.433540680	156312.9362		
5	978048	00938	956325	0067823	148707.7408		
6	1.3891643520	0.5030667850	-1.192158045	-0.433540680	95945.27543		
7	978048	568069	773245	3487385	133339.3897		
8	1.3891643520	-0.212420108	0.0428123055	-0.433540680	101164.0924		
9	978048	98732416	1.1193825	3487385	124251.9871		
10	-0.404077066	0.6223146007	1.2777826567	0.4671671043	55066.11435		
11	27285556	308288	956325	0930536	124972.5088		
12	1.3891643520	-0.093172293	1.2777826567	-1.334248465	126139.1367		
13	978048	3133023	956325	0067823	81007.19945		

kNN Regression Model

Target Column: Col6 -- Salary

Number of Neighbors: 3

Execute Cancel

The predictions will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (S)	Col5 (D)	Col6 (S)	Col7 (D)	Col8
1	1.300697775	-0.808659187	1.2777826567	84340.773109	0.0714285714	0.0357142857	0.0357142857	Entry 1
2	4581859	3574334	956325	64548	0.0714285714	0.0357142857	0.0357142857	Entry 2
3	-0.404077066	1.5762971261	0.0428123055	97791.956259	0.0714285714	0.0357142857	0.0357142857	Entry 3
4	1.3891643520	1.2185536791	1.1193825	63816	0.0714285714	0.0357142857	0.0357142857	Entry 4
5	978048	00938	956325	131644.11220	0.0714285714	0.0357142857	0.0357142857	Entry 5
6	1.3891643520	0.5030667850	-1.192158045	604865	0.0714285714	0.0357142857	0.0357142857	Entry 6
7	978048	568069	773245	158707.65206	0.0714285714	0.0357142857	0.0357142857	Entry 7
8	1.3891643520	-0.212420108	0.0428123055	62188	0.0714285714	0.0357142857	0.0357142857	Entry 8
9	978048	98732416	1.1193825	148004.25472	0.0714285714	0.0357142857	0.0357142857	Entry 9
10	-0.404077066	0.6223146007	1.2777826567	188095	0.0714285714	0.0357142857	0.0357142857	Entry 10
11	27285556	308288	956325	98210.983956	0.0714285714	0.0357142857	0.0357142857	Entry 11
12	1.3891643520	-0.093172293	1.2777826567	66666	0.0714285714	0.0357142857	0.0357142857	Entry 12
13	978048	3133023	956325	133086.50280	0.0714285714	0.0357142857	0.0357142857	Entry 13

Step 10: Validate the model

Create a new tab by pressing the "+" button on the bottom of the page with the name "VALIDATE_MODEL(.predict)".

Import data into the input spreadsheet of the "VALIDATE_MODEL(.predict)" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with options: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu bar, a workflow diagram is visible with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(fit), and VALIDATE_MODEL(predict). Below the workflow, a data table is displayed with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D), Col7, Col8, and Col9. The table contains 12 rows of data. The 'Salary' column (Col5) is highlighted in blue. The table is titled 'TRAIN_MODEL(fit)'.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
1	1.3891643520	0.450915740	0.0428123055	0.433540680	142591.2559				
2	1.300697775	1.6955449417	-1.192158045	-0.433540680	96834.67128				
3	1.3891643520	-1.285600450	-1.192158045	-0.433540680	130567.6495				
4	1.300697775	0.7415624164	1.2777826567	1.3678748889	89426.94712				
5	1.300697775	0.48505	0.956325	0.673493	104762.3985				
6	1.300697775	-0.093172293	0.06567956325	-1.334248465	75559.24411				
7	1.3891643520	1.0993058634	0.0428123055	-0.433540680	153930.9339				
8	1.3891643520	-0.212420108	-1.192158045	-0.433540680	151123.3359				
9	1.3891643520	0.4925436429	-0.927907003	-1.192158045	106019.0204				
10	1.3891643520	-1.524146081	0.0428123055	-0.433540680	142122.6673				
11	1.300697775	0.5030667850	1.2777826567	1.3678748889	95726.36719				
12	1.300697775	-0.093172293	0.06567956325	-1.334248465	66216.70162				

To validate the model browse: "Analytics" → "Existing Model Utilization". Then choose Model "(from Tab:) TRAIN_MODEL (.fit)" and transfer the "Salary" column in the output.

The screenshot shows the 'Existing Model Execution' dialog box. The 'Model' dropdown is set to '(from Tab:) TRAIN_MODEL(...)'. The 'Type' is 'kNN Model'. The 'Description' field is empty. The 'Model Input' section lists the following variables and their datatypes: Education (Double), Experience (Double), Location (Double), Job_Title (Double), Age (Double), and Gender (Double). The 'Transfer Column(s) to Output' checkbox is checked. The 'Excluded Columns' list contains: Col2 -- Education, Col3 -- Experience, Col4 -- Location, and Col5 -- Job_Title. The 'Included Columns' list contains: Col6 -- Salary. The 'Execute' button is highlighted.

The predictions will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7	Col8	Col9
1	1.3891643520	0.450915740	0.0428123055	0.433540680	142591.2559				
2	1.300697775	1.695544947	-1.192158045	0.433540680	96834.67128				
3	1.3891643520	-1.285600450	-1.192158045	0.433540680	130567.6495				
4	1.300697775	0.7415624164	1.2777826567	1.3678748889	89426.94712				
5	1.300697775	0.61346	0.95625	0.73493	104762.3985				
6	1.300697775	-0.093172293	1.2777826567	1.334248465	75559.24411				
7	1.3891643520	1.0993058634	0.0428123055	0.433540680	153930.9339				
8	1.3891643520	-0.212420108	-1.192158045	0.433540680	151123.3359				
9	1.3891643520	0.4925436429	-0.927907003	-1.192158045	106019.0204				
10	1.3891643520	-1.524146081	0.0428123055	0.433540680	142122.6673				
11	1.300697775	0.9430667850	1.2777826567	1.3678748889	95726.36719				
12	1.300697775	-0.404077066	-1.404898265	0.0428123055	66216.70162				

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (D)
1	147329.69760	Entry 318	0.0714285714	285714	Entry 5	0.1071428571	428571	142591.2559	
2	102260.31348	Entry 659	0.0357142857	1428581	Entry 15	0.1071428571	428572	96834.67128	
3	136266.16524	Entry 272	0.0	Entry 589	0.0357142857	14285685	Entry 247	0.0357142857	
4	93556.418516	Entry 208	0.0	Entry 370	0.0357142857	1428559	Entry 214	0.0357142857	
5	103811.90576	Entry 656	0.0357142857	1428564	Entry 378	0.0357142857	Entry 622	0.1071428571	
6	68721.626304	Entry 319	0.0	Entry 28	0.0357142857	1428571	Entry 342	0.0357142857	
7	142501.44185	Entry 734	0.0357142857	1428559	Entry 203	0.0714285714	Entry 521	0.1071428571	
8	133322.84315	Entry 320	0.0714285714	285714	Entry 115	0.1071428571	Entry 29	0.1071428571	
9	112748.63586	Entry 738	0.1071428571	4285708	Entry 495	0.1428571428	Entry 183	0.1785714285	
10	134140.22246	Entry 595	0.0357142857	14285685	Entry 74	0.0357142857	Entry 107	0.1071428571	
11	94956.387771	Entry 214	0.0357142857	1428581	Entry 390	0.0357142857	Entry 208	0.0714285714	
12	69531.939583	Entry 217	0.0714285714	285714	Entry 178	0.1071428571	Entry 503	0.1785714285	

Step 11: Statistics calculation

Create a new tab by pressing the "+" button on the bottom of the page with the name "STATISTICS_ACCURACIES".

Import data into the input spreadsheet of the "STATISTICS_ACCURACIES" tab from the output of the "VALIDATE_MODEL(predict)" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

User Header	Col1	Col2 (D)	Col3 (S)	Col4 (D)	Col5 (S)	Col6 (D)	Col7 (S)	Col8 (D)	Col9 (D)
1	147329.69760	Entry 318	0.0714285714	285714	Entry 5	0.1071428571	428571	142591.2559	
2	102260.31348	Entry 659	0.0357142857	1428581	Entry 15	0.1071428571	428572	96834.67128	
3	136266.16524	Entry 272	0.0	Entry 589	0.0357142857	14285685	Entry 247	0.0357142857	
4	93556.418516	Entry 208	0.0	Entry 370	0.0357142857	1428559	Entry 214	0.0357142857	
5	103811.90576	Entry 656	0.0357142857	1428564	Entry 378	0.0357142857	Entry 622	0.1071428571	
6	68721.626304	Entry 319	0.0	Entry 28	0.0357142857	1428571	Entry 342	0.0357142857	
7	142501.44185	Entry 734	0.0357142857	1428559	Entry 203	0.0714285714	Entry 521	0.1071428571	
8	133322.84315	Entry 320	0.0714285714	285714	Entry 115	0.1071428571	Entry 29	0.1071428571	
9	112748.63586	Entry 738	0.1071428571	4285708	Entry 495	0.1428571428	Entry 183	0.1785714285	
10	134140.22246	Entry 595	0.0357142857	14285685	Entry 74	0.0357142857	Entry 107	0.1071428571	
11	94956.387771	Entry 214	0.0357142857	1428581	Entry 390	0.0357142857	Entry 208	0.0714285714	
12	69531.939583	Entry 217	0.0714285714	285714	Entry 178	0.1071428571	Entry 503	0.1785714285	

Calculate the statistical metrics for the regression by browsing: "Statistics" → "Model Metrics" → "Regression Metrics".

The screenshot shows the 'Statistics' menu with options like 'Domain - APD', 'Model Metrics', 'Probability Distribution Functions', 'Descriptive Statistics', 'Confidence Intervals', 'Hypothesis Testing', 'Weight Cases', 'Random Number Generator', and 'Design of Experiments'. The 'Model Metrics' option is selected, leading to the 'Regression Statistics Metrics' dialog box.

The dialog box has two dropdown menus: 'Actual Value Column' set to 'Col9 -- Salary' and 'Prediction Value Column' set to 'Col2 -- kNN Prediction'. There are 'Execute' and 'Cancel' buttons at the bottom.

The background spreadsheet shows a table with columns: User Header, User Row ID, kNN Prediction, Closest NN1, Distance from NN1, Closest NN2, Distance from NN2, Closest NN3, Distance from NN3, and Salary. The data is for 12 users.

The results will appear on the output spreadsheet.

The screenshot shows the 'Statistics' menu with options like 'Domain - APD', 'Model Metrics', 'Probability Distribution Functions', 'Descriptive Statistics', 'Confidence Intervals', 'Hypothesis Testing', 'Weight Cases', 'Random Number Generator', and 'Design of Experiments'. The 'Model Metrics' option is selected, leading to the 'Regression Statistics Metrics' dialog box.

The dialog box has two dropdown menus: 'Actual Value Column' set to 'Col9 -- Salary' and 'Prediction Value Column' set to 'Col2 -- kNN Prediction'. There are 'Execute' and 'Cancel' buttons at the bottom.

The background spreadsheet shows a table with columns: User Header, User Row ID, kNN Prediction, Closest NN1, Distance from NN1, Closest NN2, Distance from NN2, Closest NN3, Distance from NN3, and Salary. The data is for 12 users.

Step 12: Reliability check of each record of the test set

Step 12.a: Create the domain

Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE_SALARY".

Import data into the input spreadsheet of the "EXCLUDE_SALARY" tab from the output of the "FEATURE_SELECTION_TRAIN_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from Spreadsheet".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with options: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu, a workflow diagram shows the process: IMPORT → TRAIN_TEST_SPLIT → NORMALIZE_TRAIN_SET → NORMALIZE_TEST_SET → FEATURE_SELECTION_REGRESSION → TRAIN_MODEL → VALIDATE_MODEL → STATISTICS_ACCURACIES → EXCLUDE_SALARY. Below the workflow, a data table is displayed with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6 (D). The table contains 12 rows of data. The bottom of the interface shows a tab bar with the following tabs: DRT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL, VALIDATE_MODEL, STATISTICS_ACCURACIES, EXCLUDE_SALARY.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)
1	-1.300697775	0.808659187	1.277826567	1.367874889	84620.05366	
2	4581859	3574334	956325	673493	97800.2554	
3	-0.404077066	1.5762971261	0.0428123055	1.367874889	132157.7862	
4	27285556	230034	11193825	673493	156312.9362	
5	1.3891643520	1.2185536791	1.277826567	-1.334248465	148707.7408	
6	978048	00938	956325	0067823	95945.27543	
7	1.3891643520	0.5030667850	-1.192158045	-0.433540680	133339.3897	
8	978048	568069	773245	3487385	124251.9871	
9	1.3891643520	-0.212420108	0.0428123055	-0.433540680	55066.11435	
10	978048	98732416	11193825	3487385	124972.5088	
11	-0.404077066	0.6223146007	1.277826567	0.4671671043	126139.1367	
12	27285556	308288	956325	0930536		
13	1.3891643520	-0.093172293	1.277826567	-1.334248465		
14	978048	3133023	956325	0067823		
15	-0.404077066	0.9800580477	-1.192158045	1.367874889		
16	27285556	520943	773245	673493		
17	0.4925436429	-1.643393897	1.277826567	-0.433540680		
18	1247466	0755864	956325	3487385		
19	-1.300697775	-0.450915740	1.277826567	1.367874889		
20	4581859	33536785	956325	673493		
21	1.3891643520	-1.285650450	-1.192158045	1.367874889		
22	978048	0535206	773245	673493		
23	0.4925436429	-0.927907003	0.0428123055	-0.433540680		
24	1247466	0314552	11193825	3487385		

Manipulate the data to exclude the column that corresponds to the "Salary" by browsing: "Data Transformation" → "Data Manipulation" → "Select Columns". Then select all the columns except the "Salary".

The screenshot shows the Isalos Analytics Platform interface with the 'Data Transformation' menu open. The 'Data Manipulation' option is selected, and the 'Select Column(s)' dialog box is displayed. The dialog box has two sections: 'Excluded Columns' and 'Included Columns'. The 'Excluded Columns' section contains 'Col6 -- Salary'. The 'Included Columns' section contains 'Col2 -- Education', 'Col3 -- Experience', 'Col4 -- Location', and 'Col5 -- Job_Title'. The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)
1	-1.300697775	0.808659187	1.277826567	1.367874889	84620.05366	
2	4581859	3574334	956325	673493	97800.2554	
3	-0.404077066	1.5762971261	0.0428123055	1.367874889	132157.7862	
4	27285556	230034	11193825	673493	156312.9362	
5	1.3891643520	1.2185536791	1.277826567	-1.334248465	148707.7408	
6	978048	00938	956325	0067823	95945.27543	
7	1.3891643520	0.5030667850	-1.192158045	-0.433540680	133339.3897	
8	978048	568069	773245	3487385	124251.9871	
9	1.3891643520	-0.212420108	0.0428123055	-0.433540680	55066.11435	
10	978048	98732416	11193825	3487385	124972.5088	
11	-0.404077066	0.6223146007	1.277826567	0.4671671043	126139.1367	
12	27285556	308288	956325	0930536		
13	1.3891643520	-0.093172293	1.277826567	-1.334248465		
14	978048	3133023	956325	0067823		
15	-0.404077066	0.9800580477	-1.192158045	1.367874889		
16	27285556	520943	773245	673493		
17	0.4925436429	-1.643393897	1.277826567	-0.433540680		
18	1247466	0755864	956325	3487385		
19	-1.300697775	-0.450915740	1.277826567	1.367874889		
20	4581859	33536785	956325	673493		
21	1.3891643520	-1.285650450	-1.192158045	1.367874889		
22	978048	0535206	773245	673493		
23	0.4925436429	-0.927907003	0.0428123055	-0.433540680		
24	1247466	0314552	11193825	3487385		

The results will appear on the output spreadsheet.

Create a new tab by pressing the "+" button on the bottom of the page with the name "DOMAIN".

Import data into the input spreadsheet of the "DOMAIN" tab from the output of the "EXCLUDE_SALARY" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with options: File, Edit, Data Transformation, Analytics, Statistics, Plot, Help. Below the menu bar, there is a workflow diagram with nodes: IMPORT, TRAIN_TEST_SPLIT, NORMALIZE_TRAIN_SET, NORMALIZE_TEST_SET, FEATURE_SELECTION_REGRESSION, FEATURE_SELECTION_TRAIN_SET, FEATURE_SELECTION_TEST_SET, TRAIN_MODEL(lr), VALIDATE_MODEL(predict), STATISTICS_ACCURACIES, EXCLUDE_SALARY, and DOMAIN. The DOMAIN node is highlighted. Below the workflow diagram, there is a data table with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6. The table contains 12 rows of data. The DOMAIN node is highlighted in the workflow diagram.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6
1	-1.300697775	-0.808659187	1.2777826567	1.3678748889		
2	4581859	3574334	956325	673493		
3	-0.404077066	1.5762971261	0.0428123055	1.3678748889		
4	27285556	230034	11193825	673493		
5	1.3891643520	1.2185536791	1.2777826567	-1.334248465		
6	978048	00938	956325	0067823		
7	1.3891643520	0.5030667850	-1.192158045	-0.433540680		
8	978048	568069	773245	3487385		
9	1.3891643520	-0.212420108	0.0428123055	-0.433540680		
10	978048	98732416	11193825	3487385		
11	-0.404077066	0.6223146007	1.2777826567	0.4671671043		
12	27285556	308288	956325	0930536		

Create the domain by browsing: "Statistics" → "Domain APD".

The screenshot shows the Isalos Analytics Platform interface. The 'Statistics' menu is open, showing options: Domain - APD, Model Metrics, Probability Distribution Functions, Descriptive Statistics, Confidence Intervals, Hypothesis Testing, Weight Cases, Random Number Generator, and Design of Experiments. The 'Domain - APD' option is selected. Below the menu, there is a data table with columns: User Header, Col1, Col2 (D), Col3 (D), Col4 (D), Col5 (D), Col6. The table contains 12 rows of data. The 'Domain - APD' dialog box is open, showing the formula: $APD = d + Z\sigma$, with $Z = 0.5$. The 'Perform Computations' dropdown is set to 'CPU (double precision)'. The 'Execute' button is highlighted.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6
1	-1.300697775	-0.808659187	1.2777826567	1.3678748889		
2	4581859	3574334	956325	673493		
3	-0.404077066	1.5762971261	0.0428123055	1.3678748889		
4	27285556	230034	11193825	673493		
5	1.3891643520	1.2185536791	1.2777826567	-1.334248465		
6	978048	00938	956325	0067823		
7	1.3891643520	0.5030667850	-1.192158045	-0.433540680		
8	978048	568069	773245	3487385		
9	1.3891643520	-0.212420108	0.0428123055	-0.433540680		
10	978048	98732416	11193825	3487385		
11	-0.404077066	0.6223146007	1.2777826567	0.4671671043		
12	27285556	308288	956325	0930536		

Domain - APD

$APD = d + Z\sigma$, $Z = 0.5$

Perform Computations CPU (double precision)

Execute Cancel

The results will appear on the output spreadsheet.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7	Col8	Col9
1	1.300697775	-0.808659187	1.277826567	1.3678748889	673493				
2	-0.404077066	1.5762971261	0.0428123055	1.3678748889	673493				
3	1.3891643520	1.2185536791	1.277826567	1.334248465	0067823				
4	1.3891643520	0.5030667850	-1.192158045	-0.433540680	3487385				
5	1.3891643520	-0.212420108	0.0428123055	-0.433540680	3487385				
6	-0.404077066	0.6223146007	1.277826567	0.4071671043	0930536				
7	1.3891643520	2.2913113023	1.277826567	-1.334248465	0067823				
8	-0.404077066	0.9800580477	-1.192158045	1.3678748889	673493				
9	0.4925436429	-1.643393897	1.277826567	-0.433540680	3487385				
10	-1.300697775	-0.450915740	1.277826567	1.3678748889	673493				
11	1.3891643520	-1.285604500	-1.192158045	1.3678748889	673493				
12	1.3891643520	0.535206	773245	0.0428123055	-0.433540680				

Step 12.b: Check the test set reliability

Create a new tab by pressing the "+" button on the bottom of the page with the name "EXCLUDE_SALARY_TEST_SET".

Import data into the input spreadsheet of the "EXCLUDE_SALARY_TEST_SET" tab from the output of the "FEATURE_SELECTION_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7	Col8	Col9
1	1.3891643520	-0.450915740	0.0428123055	-0.433540680	142591.2559				
2	1.3891643520	33526785	11193825	3487385	96834.67128				
3	-1.300697775	1.6955449417	-1.192158045	-0.433540680	130567.6495				
4	1.3891643520	-1.285604500	-1.192158045	-0.433540680	89426.94712				
5	1.3891643520	0.535206	773245	0.0428123055	104762.3985				
6	-1.300697775	-0.331667924	1.277826567	1.3678748889	75559.24411				
7	1.3891643520	0.9800580477	-1.192158045	1.3678748889	153930.9339				
8	0.4925436429	-1.643393897	1.277826567	-0.433540680	151123.3359				
9	1.3891643520	0.535206	773245	0.0428123055	106019.0204				
10	1.3891643520	-1.524146081	0.0428123055	-0.433540680	142122.6673				
11	1.3891643520	4015644	11193825	3487385	95726.36719				
12	-0.404077066	-1.404898265	0.0428123055	-1.334248465	66216.70162				

Filter the data to exclude the column that corresponds to the "Salary" by browsing: "Data Transformation" → "Data Manipulation" → "Select Columns". Then select all the columns except "Salary".

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6 (D)	Col7
1	1.3891643520	-0.450915740	0.0428123055	-0.433540680	142591.2559		
2	1.300697775	1.8955449417	-1.192158045	-0.433540680	96834.67128		
3	1.3891643520	-1.285650450	-1.192158045	-0.433540680	130567.6495		
4	-1.300697775	0.7415624164	1.2777826567	1.3678748889	89426.94712		
5	-0.404077066	-0.331667924	0.0428123055	-0.433540680	104762.3985		
6	-1.300697775	-0.093172293	1.2777826567	-1.334248465	75559.24411		
7	1.3891643520	0.093058634	0.0428123055	-0.433540680	153930.9339		
8	1.3891643520	-0.212420108	-1.192158045	-0.433540680	151123.3559		
9	0.4925436429	-0.927907003	-1.192158045	-0.433540680	106019.0204		
10	1.3891643520	-1.524146081	0.0428123055	-0.433540680	142122.6673		
11	-1.300697775	0.5030667850	1.2777826567	1.3678748889	95726.36719		
12	-0.404077066	-1.404898265	0.0428123055	-1.334248465	66216.70162		

The results will appear on the output spreadsheet.

Create a new tab by pressing the "+" button on the bottom of the page with the name "RELIABILITY".

Import data into the input spreadsheet of the "RELIABILITY" tab from the output of the "EXCLUDE_SALARY_TEST_SET" tab by right-clicking on the input spreadsheet and then choosing "Import from SpreadSheet".

User Header	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1.3891643520	-0.450915740	0.0428123055	-0.433540680	142591.2559			
2	1.300697775	1.8955449417	-1.192158045	-0.433540680	96834.67128			
3	1.3891643520	-1.285650450	-1.192158045	-0.433540680	130567.6495			
4	-1.300697775	0.7415624164	1.2777826567	1.3678748889	89426.94712			
5	-0.404077066	-0.331667924	0.0428123055	-0.433540680	104762.3985			
6	-1.300697775	-0.093172293	1.2777826567	-1.334248465	75559.24411			
7	1.3891643520	0.093058634	0.0428123055	-0.433540680	153930.9339			
8	1.3891643520	-0.212420108	-1.192158045	-0.433540680	151123.3559			
9	0.4925436429	-0.927907003	-1.192158045	-0.433540680	106019.0204			
10	1.3891643520	-1.524146081	0.0428123055	-0.433540680	142122.6673			
11	-1.300697775	0.5030667850	1.2777826567	1.3678748889	95726.36719			
12	-0.404077066	-1.404898265	0.0428123055	-1.334248465	66216.70162			

Check the Reliability by browsing: "Analytics" → "Existing Model Utilization". Then select as Model "(from Tab:) DOMAIN".

The screenshot shows the Isalos Analytics Platform interface. At the top, there is a menu bar with 'File', 'Edit', 'Data Transformation', 'Analytics', 'Statistics', 'Plot', and 'Help'. Below the menu is a workflow diagram with nodes: 'IMPORT', 'TRAIN_TEST_SPLIT', 'NORMALIZE_TRAIN_SET', 'FEATURE_SELECTION_REGRESSION', 'NORMALIZE_TEST_SET', 'FEATURE_SELECTION_TRAIN_SET', 'EXCLUDE_SALARY', and 'DOMAIN'. Below the workflow is a data table with 7 columns: 'User Header', 'Col1', 'Col2 (D)', 'Col3 (D)', 'Col4 (D)', 'Col5 (D)', 'Col6', and 'Col7'. The table contains 12 rows of data.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1	1.3891643520	-0.450915740	0.0428123055	-0.433540680			
2	1.3891643520	-1.300697775	1.6955449417	-1.192158045	-0.433540680		
3	1.3891643520	4581859	970254	773245	3487385		
4	1.3891643520	4581859	048505	956325	673493		
5	1.3891643520	-0.404077066	-0.331667924	1.2777826567	1.3678748889		
6	1.3891643520	27285556	661346	956325	673493		
7	1.3891643520	-1.300697775	-0.093172293	1.2777826567	1.334248465		
8	1.3891643520	4581859	3133023	956325	0067823		
9	1.3891643520	1.0993058634	0.0428123055	-0.433540680			
10	1.3891643520	978048	269162	11193825	3487385		
11	1.3891643520	-0.212420108	-1.192158045	-0.433540680			
12	1.3891643520	978048	98732416	773245	3487385		

The 'Existing Model Execution' dialog box is shown. It has a 'Model' dropdown set to '(from Tab:)DOMAIN' and a 'Type' dropdown set to 'APD Model'. Below these is a 'Description' text area. Under 'Model Input', there is a list of variables and their datatypes: Header -> Datatype, Education -> Double, Experience -> Double, Location -> Double, Job_Title -> Double, Age -> Double, and Gender -> Double. At the bottom, there is a checkbox 'Transfer Column(s) to Output' which is unchecked. There are 'Execute' and 'Cancel' buttons at the bottom right.

The results will appear on the output spreadsheet.

The screenshot shows the Isalos Analytics Platform interface with the workflow diagram and the output spreadsheet. The workflow diagram is the same as in the first screenshot. The output spreadsheet has 8 columns: 'User Header', 'Col1', 'Col2 (D)', 'Col3 (D)', 'Col4 (D)', 'Col5 (D)', 'Col6', and 'Col7'. The table contains 12 rows of data, including the 'Domain' and 'APD' columns.

User Header	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)	Col6	Col7
1	1.3891643520	-0.450915740	0.0428123055	-0.433540680			
2	1.3891643520	-1.300697775	1.6955449417	-1.192158045	-0.433540680		
3	1.3891643520	4581859	970254	773245	3487385		
4	1.3891643520	4581859	048505	956325	673493		
5	1.3891643520	-0.404077066	-0.331667924	1.2777826567	1.3678748889		
6	1.3891643520	27285556	661346	956325	673493		
7	1.3891643520	-1.300697775	-0.093172293	1.2777826567	1.334248465		
8	1.3891643520	4581859	3133023	956325	0067823		
9	1.3891643520	1.0993058634	0.0428123055	-0.433540680			
10	1.3891643520	978048	269162	11193825	3487385		
11	1.3891643520	-0.212420108	-1.192158045	-0.433540680			
12	1.3891643520	978048	98732416	773245	3487385		

There are no unreliable samples in the test set.

Final Isalos Workflow

Following the above-described steps, the final workflow on Isalos will look like this:

