**Power BI vs JMP tool:**

**Power BI**: It is a great tool for Data Visualization. Capable of building sophisticated data models with multiple related tables. Reports that are generated are very difficult to share with others. A lot is possible with custom calculated columns using DAX but the software is focused on visuals; exploration such as outliers, finding correlation etc., are manual calculations.

**JMP:**

JMP is a complete tool. It Includes all functionality that are needed for statistics, data mining and visualization within the same software interface. I found that most functions I could learn just by clicking on the menus.

It is not necessary to use multiple applications to complete a data analysis.

**Assumptions:**

The following table represents the number of clicks needed to perform basic tasks in each of the software applications.

|  |  |  |
| --- | --- | --- |
| **Tasks** | **Clicks/Keystrokes** | |
|  | **Power BI** | **JMP** |
| Import CSV | 4 | 3 |
| Bar Graph | 3 | 4 |
| Scatter Plot | 3 | 3 |
| Descriptive Statistics | 19 | 3 |
| Linear Regression | 10 | 5 |

Linear Regression counts are for scatter plots with trendlines added (simple linear regressions).

Feature Offerings:

|  |  |  |
| --- | --- | --- |
| Dashboarding | Data Exploration | Data Cleaning |
| Ability to combine graphs/tables into one place that can be updated automatically when the data changes. | Utilities to quickly understand a data set such as outlier, missing value and duplicate values identification. | Utilities to transform data with functions such as combining tables, replace missing values, remove outliers or duplicates or change column/row structures. |
| 0 | Does not have the feature | |
| 1 | Feature designed | |

|  |  |  |
| --- | --- | --- |
| Task | Feature Offerings | |
|  | Power BI | JMP |
| Dashboarding | 1 | 1 |
| Data Exploration | 0 | 1 |
| Data Cleaning | 0 | 1 |

Using Python and DAX functions we can explore the data:

Once we have imported the data into a Pandas data frame, we can use the below methods to get a sense of what the data looks like:

|  |  |  |
| --- | --- | --- |
| **Functions** | **Description** | **DAX/ Python Script** |
| df.shape() | Prints number of rows and columns in dataframe | Python |
| df.head(n) | Prints first n rows of the DataFrame | Python |
| df.tail(n) | Prints last n rows of the DataFrame | Python |
| df.info() | Index, Datatype and Memory information | Python |
| df.describe() | Summary statistics for numerical columns | Python |
| s.value\_counts(dropna=False) | Views unique values and counts | Python |
| df.apply(pd.Series.value\_counts) | Unique values and counts for all columns | Python |
| df.mean() | Returns the mean of all columns | Python |
| df.corr() | Returns the correlation between columns in a DataFrame | Python |
| df.count() | Returns the number of non-null values in each DataFrame column | Python |
| df.max() | Returns the highest value in each column | Python |
| df.min() | Returns the lowest value in each column | Python |
| df.median() | Returns the median of each column | Python |
| df.std() | Returns the standard deviation of each column | Python |

|  |  |  |
| --- | --- | --- |
| **Statistical Functions in Power BI** | **Description** | **DAX/Python** |
| ADDCOLUMNS | Adds calculated columns to the given table or table expression | DAX |
| APPROXIMATEDISTINCTCOUNT | Returns the approximate number of rows that contain distinct values in a column. | DAX |
| AVERAGE | Returns the average (arithmetic mean) of all the numbers in a column | DAX |
| COUNT | The COUNT function counts the number of cells in a column that contain numbers. | DAX |
| COUNTBLANK | Counts the number of blank cells in a column. | DAX |
| COUNTROWS | The COUNTROWS function counts the number of rows in the specified table, or in a table defined by an expression. | DAX |
| DISTINCTCOUNT | Counts the number of distinct values in a column. | DAX |
| GEOMEAN | Returns the geometric mean of the numbers in a column. | DAX |
| MAX | Returns the largest numeric value in a column, or between two scalar expressions | DAX |
| MEDIAN | Returns the median of numbers in a column. | DAX |
| MIN | Returns the smallest numeric value in a column, or between two scalar expressions. | DAX |
| NORM.DIST | Returns the normal distribution for the specified mean and standard deviation | DAX |
| POISSON.DIST | Returns the Poisson distribution. | DAX |
| ROW | Returns a table with a single row containing values that result from the expressions given to each column. | DAX |
| SAMPLE | Returns a sample of N rows from the specified table. | DAX |
| SELECTCOLUMNS | Adds calculated columns to the given table or table expression. | DAX |
| STDEV.P | Returns the standard deviation of the entire population | DAX |
| SUMMARIZE | Returns a summary table for the requested totals over a set of groups. | DAX |
| Correlation coefficient | To know the correlation between variables | DAX & inbuilt feature |
| Outliers | To remove distant observations | DAX |