Excel Data Cleaning Project

Data cleaning is key aspect of data analysis. Data analyst spend most of the time in data cleaning.

Data cleaning is the process of removing inconsistent data and making prepared for data analysis.

Data cleaning process :

* Check for empty records and delet all the Blank records by applying filters.
* Remove duplicates by using conditional formatting & Advance filtering.
* Identify outliers by using Visualization method or statical method.

For visualization you can use scatter plot and histogram.

After detecting outliers delete total row if the data set have more records.

* Identify missing values by using count( ).

If the missing values in more than 30% then delete column.

If the missing values in more than 50% then delete column.

Missing values can be treated by using data imputation techniques Mean, median for continues data and mode for categorical data.

* Change data type or convert data type, by using find & replace, text to column

Example : Date format, Currency, Location etc.

* For inconsistency data use data validation technique (Based on criteria, circle validation),sort & filter.

Data set : Data set is consist of 70 records of sales data. After completing data cleaning process it reached to 60 records

In this project

* For empty rows first i removed empty rows by applying filters.
* For outliers detection I used scatter plot for Age column. After identifying outliers I deleted entire row.
* For missing values I identified missing values by using count( ) for all rows after identifying missing values I filled missing values by using data imputation technique mean for age column and rating column.
* Converted date type into date format (dd-mm-yyyy).
* Converted data type of currencies in structured format.
* Removed data duplicates by using conditional formatting.

Before Vs After data cleaning

|  |  |  |
| --- | --- | --- |
|  | Before Data cleaning | After Data cleaning |
| Total rows | 70 | 60 |
| Data duplicates | 8 | 0 |
| Missing values(Age) | 4 | 0 |
| Data type (Currencies) | $,NZ$, ₽ ,₺ | $ |
| Data type (Date) | MM-DD-YYYY | DD-MM-YYYY |
| Missing values(Ratings) | 3 | 0 |
| Outliers | 1 | 0 |

By

katari Pavan

9866645774

Pavankatari24@gmail.com