

Data Cleaning Process Documentation

The cleaning process for Excel Workbooks

Excel workbooks chosen for Analysis

1. Q4 December 2021
2. Q1 March 2022
3. Q2 June 2022
4. Q3 September 2022

Cleaning Blanks

- 1 select all columns with data in one spreadsheet.
2. Go to Find and Select on the Home menu in Excel.
3. Find 'Go to Special'
4. select 'Blanks'
5. Once cells are selected, choose "delete sheet rows" from the home tab.
6. save the file in the recommended name.
7. Repeat for the rest of the Excel documents.
8. for larger Excel documents, sort and then filter by column by blanks and delete those rows with empty records.

Calculating ride length

1. Create a new column between ended_at column and start_station_name column and name it ride_length. It is now column E.
2. Insert formula at E2: =(D2-C2)
3. Copy the formula throughout the rest of the column cells by selecting Fill and selecting the down arrow button.
4. Change the time format to HH:MM:SS by going to the Number menu on the Home Tab.

Calculating day of week

5. Create a new column F, name it day_of_the_week and calculate the day of the week using the formula: =WEEKDAY(C2,1). Format is as a General with no decimals.
6. NB: 1 stand for Sunday and the number 7 for Saturday.
7. Remove duplicates if any.
8. Convert ride_length time from HH:MM:SS format by using the Number menu on the home tab. Select time and the format 0:00 for minutes.

Create Pivot Tables

Pivot tables can be used for calculations. Please refer to the pdf :

2022_09-Seasonal_Analysis_Summary

Create Graphs for each month mentioned above which will be used in PowerPoint

RESULT: data is clean and ready to analyze