Data Intake Report

Name: Final Cab Data
Report date: 07.03.2021
Internship Batch:LISP01
Version:1.0
Data intake by:Ines Perko
Data intake reviewer:<intern who reviewed the report>
Data storage location: https://github.com/inesp93/Cab_Data_Project

Tabular data details:

Total number of observations	359392
Total number of files	1
Total number of features	12
Base format of the file	.csv
Size of the data	29,9 MB

Proposed Approach:

- Date of Travel in Cab_Data.csv changed from 5 digit number format to date format in Excel.
- Removed duplicates & empty rows (but in each case no duplicates or empty rows)
- Left join of the Transactions_ID on the Cab_Data, then left join of the Customer_ID
- Cleaned empty rows, but there is none
- File too big to upload to GitHub

Tabular data details: (Hypothesis one)

Total number of observations	2
Total number of files	4
Total number of features	4
Base format of the file	.csv
Size of the data	354 bytes

Proposed Approach:

- Hypothesis one: Yellow Cab has a higher profit per ride than the Pink Cab
- Extract the number of the total profit
- Extract the number of rides
- Calculate the profit per ride
- Hypothesis is correct

Tabular data details: (Hypothesis two)

Total number of observations	20
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.37 KB

Proposed Approach:

- Hypothesis two: Yellow Cab Company has a majority of Profit Per Capita in a different city
- Extract the number of the profit in cities
- Calculate Profit Per Capita
- Comparing the values
- Hypothesis is correct
- The question arises, does the Yellow Cab Company has a higher user ratio in the cities

Tabular data details: (Hypothesis three)

Total number of observations	3
Total number of files	3
Total number of features	9
Base format of the file	.csv
Size of the data	479 Bytes

Proposed Approach:

- Hypothesis three: Yellow Cab Company has a higher number of both male and female users than the Pink Cab Company, and the highest number of users are the male customers aged (30-60) of the Yellow Cab Company
- Extract the number of the male and female customers with respect to the company
- Extract the number of customers with respect to the age
- Hypothesis is correct

Tabular data details: (Hypothesis four)

Total number of observations	3
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	151 Bytes

Proposed Approach:

- Hypothesis four: Card Payment is the most used
- Extract the number of the male and female users who use card payment method and cash payment method
- Hypothesis is correct

Tabular data details: (Hypothesis five)

Total number of observations	3
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	102 Bytes

Proposed Approach:

- Hypothesis five: The customers are usually from the middle or high income class
- Separate customers based on their income
- Hypothesis is correct