

Explore_bikeshare_data

August 5, 2024

0.0.1 Explore Bike Share Data

For this project, your goal is to ask and answer three questions about the available bikeshare data from Washington, Chicago, and New York. This notebook can be submitted directly through the workspace when you are confident in your results.

You will be graded against the project [Rubric](#) by a mentor after you have submitted. To get you started, you can use the template below, but feel free to be creative in your solutions!

Caricamento CSV Carico i csv nei rispettivi dataframe

```
[2]: ny = read.csv('new-york-city.csv')
wash = read.csv('washington.csv')
chi = read.csv('chicago.csv')
```

Analisi Dati e Schema Effettuo un'analisi preliminare dei 3 dataset di partenza utilizzando la funzione summary.

New York

```
[36]: summary(ny)
```

| X | Start.Time | End.Time | Trip.Duration |
|-----------------|------------------|------------------|-----------------|
| Min. : 33 | Length:300000 | Length:300000 | Min. : 61.0 |
| 1st Qu.:1707416 | Class :character | Class :character | 1st Qu.: 368.0 |
| Median :3405756 | Mode :character | Mode :character | Median : 609.0 |
| Mean :3407026 | | | Mean : 899.7 |
| 3rd Qu.:5108762 | | | 3rd Qu.: 1054.0 |
| Max. :6816152 | | | Max. :2155775.0 |

| Start.Station | End.Station | User.Type | Gender |
|------------------|------------------|------------------|------------------|
| Length:300000 | Length:300000 | Length:300000 | Length:300000 |
| Class :character | Class :character | Class :character | Class :character |
| Mode :character | Mode :character | Mode :character | Mode :character |

| Birth.Year |
|------------|
| Min. :1885 |

```

1st Qu.:1970
Median :1981
Mean   :1978
3rd Qu.:1988
Max.   :2001
NA's   :28220

```

Washington

```
[37]: summary(wash)
```

```

      X      Start.Time      End.Time      Trip.Duration
Min.   :      7  Length:300000  Length:300000  Min.    :   60.0
1st Qu.: 436394  Class :character  Class :character  1st Qu.:  410.6
Median : 875064  Mode  :character  Mode  :character  Median :   706.5
Mean   : 875404                                     Mean   :  1237.3
3rd Qu.:1313148                                     3rd Qu.:  1229.4
Max.   :1751446                                     Max.   :1235662.2

Start.Station  End.Station  User.Type
Length:300000  Length:300000  Length:300000
Class :character  Class :character  Class :character
Mode  :character  Mode  :character  Mode  :character

```

Chicago

```
[38]: summary(chi)
```

```

      X      Start.Time      End.Time      Trip.Duration
Min.   :      4  Length:300000  Length:300000  Min.    :   60.0
1st Qu.: 387137  Class :character  Class :character  1st Qu.:  393.0
Median : 777104  Mode  :character  Mode  :character  Median :   670.0
Mean   : 776346                                     Mean   :   936.2
3rd Qu.:1164065                                     3rd Qu.:  1125.0
Max.   :1551500                                     Max.   :  86224.0

Start.Station  End.Station  User.Type      Gender
Length:300000  Length:300000  Length:300000  Length:300000
Class :character  Class :character  Class :character  Class :character
Mode  :character  Mode  :character  Mode  :character  Mode  :character

```

```

Birth.Year
Min.   :1899
1st Qu.:1975

```

```

Median :1984
Mean   :1981
3rd Qu.:1989
Max.    :2016
NA's    :61019

```

Da questa analisi si evincono i seguenti punti: - New York: La variabile Birth.Year ha 28,220 valori mancanti. - Chicago: La variabile Birth.Year ha 61,019 valori mancanti, che rappresentano una percentuale significativa del totale. - Washington: Mancano completamente le informazioni su Gender e Birth.Year, limitando l'analisi ai soli dati sui viaggi e alle stazioni.

Data Cleaning e costruzione del dataset finale Di seguito il codice che userò per mergiare i 3 dataframe uniformando ed omologando i campi per ottenere un dataset finale, comprensivo di tutte le informazioni essenziali.

New York cleaning and standardization

```

[80]: names(ny)[names(ny) == "X"] <- "Trip.id"
ny$Birth.Year <- as.integer(ny$Birth.Year)
ny$City <- rep('New York', times = nrow(ny))

```

Washington cleaning and standardization

```

[81]: names(wash)[names(wash) == "X"] <- "Trip.id"
wash$Trip.Duration <- as.integer(wash$Trip.Duration)
wash$Gender <- as.character(rep(NA, times = nrow(wash)))
wash$Birth.Year <- as.integer(rep(NA, times = nrow(wash)))
wash$City <- rep('Washington', times = nrow(wash))

```

Chicago cleaning and standardization

```

[82]: names(chi)[names(chi) == "X"] <- "Trip.id"
chi$Birth.Year <- as.integer(chi$Birth.Year)
chi$City <- rep('Chicago', times = nrow(chi))

```

```

[83]: final_ds <- data.frame()
final_ds <- rbind(ny, wash, chi)
summary(final_ds)

```

| Trip.id | Start.Time | End.Time | Trip.Duration |
|-----------------|------------------|------------------|---------------|
| Min. : 4 | Length:900000 | Length:900000 | Min. : 60 |
| 1st Qu.: 551291 | Class :character | Class :character | 1st Qu.: 389 |
| Median :1102220 | Mode :character | Mode :character | Median : 660 |
| Mean :1686259 | | | Mean : 1024 |
| 3rd Qu.:1742392 | | | 3rd Qu.: 1135 |
| Max. :6816152 | | | Max. :2155775 |

| Start.Station | End.Station | User.Type | Gender |
|---------------|---------------|---------------|---------------|
| Length:900000 | Length:900000 | Length:900000 | Length:900000 |

```
Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
Birth.Year      City
Min.   :1885     Length:900000
1st Qu.:1972     Class :character
Median :1983     Mode  :character
Mean   :1980
3rd Qu.:1988
Max.   :2016
NA's   :389239
```

0.0.2 Question 1

Dividendo per fasce orarie di 3 ore la giornata, qual'è la fascia in cui vi sono più viaggi? c'è differenza tra le città?

[]:

Summary of your question 1 results goes here.

0.0.3 Question 2

Il volume di viaggi varia se il giorno è feriale o festivo?*

[]:

Summary of your question 2 results goes here.

0.0.4 Question 3

Il target di driver diviso per fascia d'età e sesso sui dati disponibili dove si incentra?

[]:

Summary of your question 2 results goes here.

0.1 Finishing Up

Congratulations! You have reached the end of the Explore Bikeshare Data Project. You should be very proud of all you have accomplished!

Tip: Once you are satisfied with your work here, check over your report to make sure that it satisfies all the areas of the [rubric](#).

0.2 Directions to Submit

Before you submit your project, you need to create a .html or .pdf version of this notebook in the workspace here. To do that, run the code cell below. If it worked correctly, you should get a return code of 0, and you should see the generated .html file in the workspace directory (click on the orange Jupyter icon in the upper left).

Alternatively, you can download this report as .html via the **File > Download as** submenu, and then manually upload it into the workspace directory by clicking on the orange Jupyter icon in the upper left, then using the Upload button.

Once you’ve done this, you can submit your project by clicking on the “Submit Project” button in the lower right here. This will create and submit a zip file with this .ipynb doc and the .html or .pdf version you created. Congratulations!

```
[72]: system('python -m nbconvert --to pdf Explore_bikeshare_data.ipynb')
```