

Project: Communicate Data Findings

Table of Contents

Introduction

- [Data Source](#)
- [Libraries](#)

Data Wrangling

- [Data Gathering](#)
- [Data Assessing](#)
- [Data Cleaning](#)

Exploratory Visualization

- [Univariate Exploration](#)
- [Bivariate Exploration](#)
- [Multivariate Exploration](#)

Introduction

This project contains [Exploration_visualization.ipynb](#) which contains sections of the exploration, starting from loading in the data, working through univariate visualizations, and ending with bivariate and multivariate exploration; and [Explanatory_slide_deck.ipynb](#) which contains the visualizations specially for slide deck and findings.

1. Data Source

- **Ford GoBike System Data:**

This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area ([Link](#)).

- **Data Range:**

Since the data after 04/2019 is showing as baywheels data and in different format, so we will request the data from \ from 04/2018 - 04/2019.

2. Libraries

```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from zipfile import ZipFile
import requests
import glob
from pandas.api.types import is_datetime64_any_dtype as is_datetime
from sklearn.cluster import KMeans
```

```
import calendar


%config InlineBackend.figure_format = 'retina'
%matplotlib inline
```


Data Wrangling


- Request the [data](#) via Requests from 04/2018 - 04/2019.
- Assess the data in the aspect to Quality and Tideness.
- Clean the dataset before visualization.


Exploratory Visualization


- Here are some key findings. **How long are people riding these bikes?**

 In this analysis we'd like to see the distribution of people riding bikes in secs. Since the distribution is right skewed here we set the top limit of 2200 secs to remove some outliers with extremely long riding time. We can see tha the peak useseage falls in 250-740 secs which is around 4.5mins, so generally speaking, people like to use gobike for short routes.

How long are people riding these bikes?  We can see from the plot that Tuesday to Friday has the most trips taken, followed by Monday and Friday. Trips in Saturday and Sunday are approximately half of the rides than in weekdays.

When are most trips taken in terms of time of the day?  The most frequent starting hours are at commute peak time 8 and at 17. Lets see if this pattern is specially for weekday.

Most Ride in time of day and week of day  We can see that oveall people turns to take rides more on Weekday commute time (16-20pm and 8:12am). This pattern is not that obvious during weekend.

Does user have different ride habit of in case of the riding duration in different month?  We can see that most compared to non-subscribed customer group, the subscribed customer group turns to have ride in slight difference of each day of week and the average ride time is lower.