

Analyzing Chicago's Bike Share Data

Abdulaziz Abdullah Binlaksar

Abstract

Over the past decade, bicycle-sharing systems have been growing in number and popularity in cities across the world. Bicycle-sharing systems allow users to rent bicycles on a very short-term basis for a price. This allows people to borrow a bike from point A and return it at point B, though they can also return it to the same location if they'd like to just go for a ride. Regardless, each bike can serve several users per day.

In this project, the aim is to understand the patterns of renting bicycles in Chicago city by using a real data provided by divvy bicycle-sharing company and to provide actionable insights to the company. understanding the patterns and trends of bicycle-sharing system in Chicago could help the company to make data-driven decisions, which could increase their profits and help the stakeholders to find potential growths opportunities.

Data

Bike sharing data is provided by davvy company in Chicago city. The duration of data is in the fourth quarter of 2019 and it consists of 704054 record and 12 variables before any kind of data cleaning or feature engineering.

The data source:

<https://divvy-tripdata.s3.amazonaws.com/index.html>

Tools:

Because of the huge size of the dataset, the use of pandas package in python is essential to deal with the data from importing, assessing the data for any quality or tidiness issues, cleaning the data, the exploratory data analysis, sharing the insights that we will find and eventually using the data to predict future outcomes. Other packages will be used in this project like numpy, matplotlib, seaborn and sklearn. In the next paragraph the workflow and tools that will be used in the project is discussed.

Algorithm and workflow:

1. Data gathering: The process of importing the data into jupyter notebook, the package that will be used for this step is pandas.
2. Data assessing: The process of finding any issues in the data that could disrupt our analysis, pandas will also be used for this purpose.
3. Data cleaning: The process of cleaning the data from the issues that we found during the previous step, pandas will also be used for this purpose.
4. Feature engineering: The process of making new column from existing columns that may land a better insight that could be used to enhance our analysis, and the tool for this task is pandas.
5. Exploratory Data Analysis (EDA): At this step we use different statistical measures and different types of visualizations to explore and investigate our data to find or spot the patterns and trends in it. After this step the findings will be presented in a clearer way in a form known as explanatory analysis, here we will use different tools packages like matplotlib, numpy, seaborn and plotly.
6. Model Building and Prediction: Here comes the final part of the project, at this step we have explained the explored or data through different insightful and informative charts for the past data, but we want to get advantage of machine learning to predict what could happen in the future based in our data. There are many tools to use at this step, it will be either scipy, or sklearn.