TLC Trip Record Data

Yellow Taxi

Introduction

The New York City Taxi and Limousine Commission (TLC), Over 200,000 TLC licensees complete approximately 1,000,000 trips each day. According to TLC the data it is recorded since 2009 - 2021

preprocess

Purpose of the project

1. Predict the fare amount of the ride.
2. Who effects on the fare amount
3. Visiualize the data
4. Visiualize the features

Chosen year and month

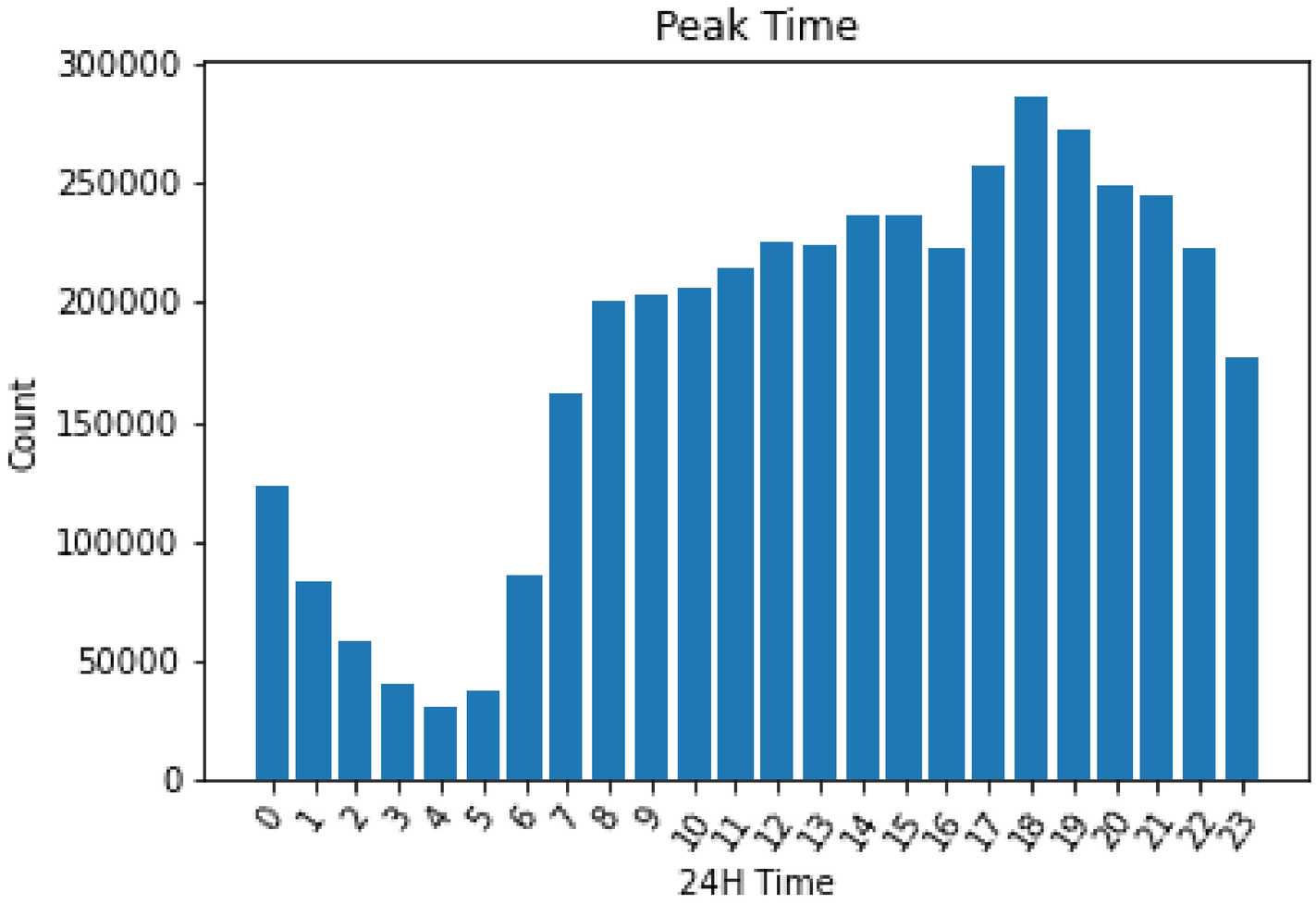
In this project the prediction and visualization will be on the dataset of October 2019

Who effects on the target?

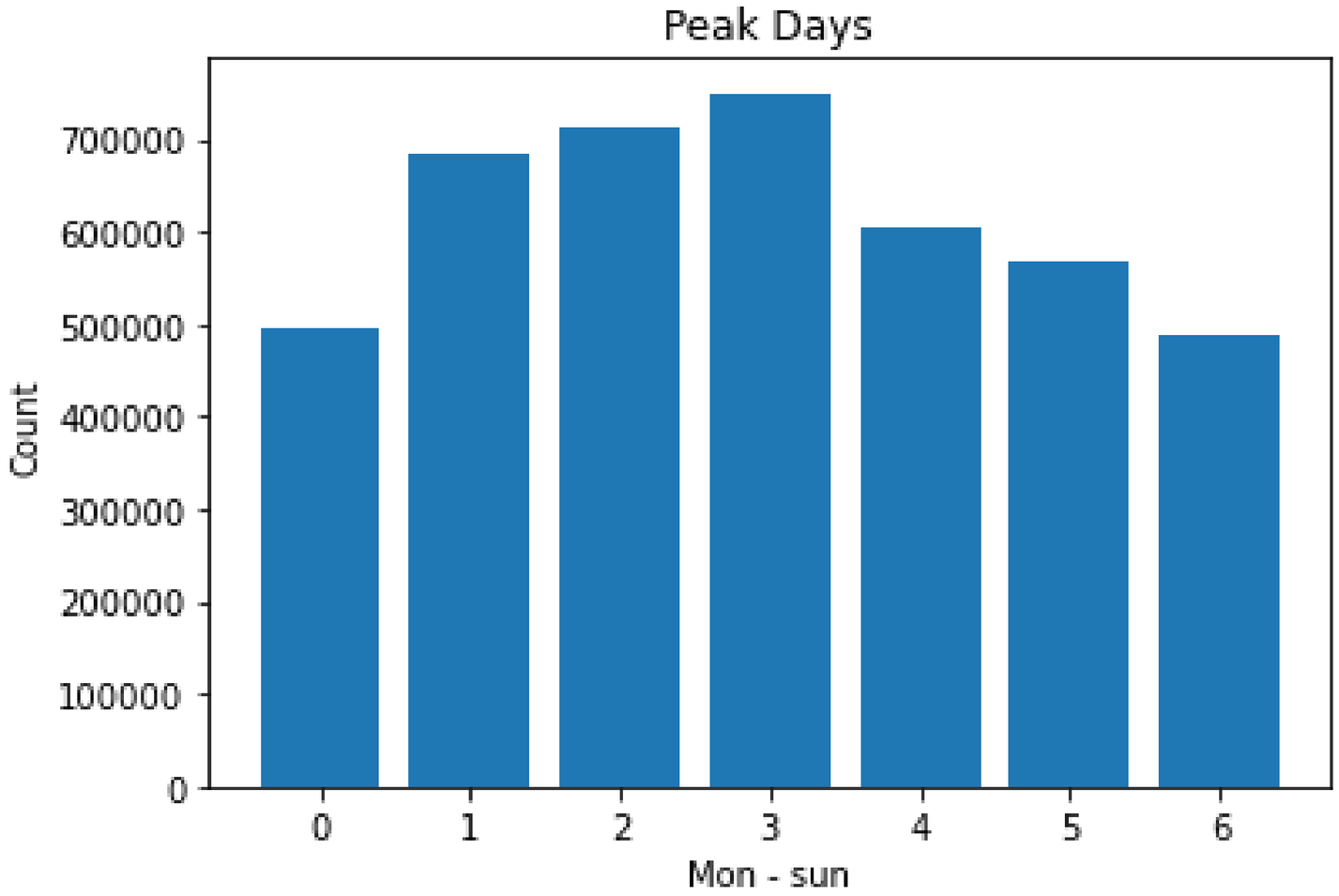
1. Distance
2. Taxi car size
3. Peak hours
4. Rate code id
5. Peak days
6. Duration

Correlation between distance and fare amount

Whata are the peak hours?



Whata are the peak days?



Data model

The Followed benchmarks in choosing the best model

Baseline feature set:
~.874 R^2

Add Category features (RatecodeID):
~.894 R^2

Add Category features (VendorID):
~.894 R^2

Add Several polynomial transforms:
~.897 R^2

Add Several interaction terms:
~.911 R^2

Add Standard-scaling Features terms:
~.911 R^2