Identifying factors favorable to demand for bike rentals

Project Proposal by: Himani Kaushik

Abstract

Every business is experiencing financial crunch because of the pandemic. Hence, to decrease the risk, prediction of important factors that affect the business and using the same factors to design a new business strategy is essential for business owners. BestBikes, a bike rental and sharing provider, also plans to analyze the demand for shared bikes and understand the factors affecting the demand. The dataset utilized for the analysis consists of daily count of rental bikes and several other features including temperature, weather, windspeed, seasons, etc. The impact hypothesis **'Higher temperature and less windspeed will be favorable to the demand for bike rentals'** connects the analysis with the business plan.

Design

The project is designed to explore a bike rental dataset and present business insights to the client. The data is utilized to build and select the most accurate predictive model for bike rentals based on several factors. Regression model, through python, is implemented to discover these significant factors that can be used to design a new strategy. Further, google spreadsheets and tableau are used to provide meaningful and interactive insights to the client.

Data

The data is downloaded from UCI repository website. Several features like temperature, windspeed, count, registered users, humidity, etc are numeric features. However, there are few categorical features as well like seasons, weather, weekdays, months, etc. These categorical features are converted to dummy variables. The count of rentals is a continuous, numeric variable and can also be considered an important KPI. Hence, it is designated as a predictive variable and other features are used to predict the demand for bike rentals.

Algorithms

- Exploratory Data Analysis in python, google spreadsheets and tableau
- Aggregation and visualization in google spreadsheets and tableau
- Interactive dashboards in Tableau
- Feature Engineering:
    - Categorical features were converted to dummy variables.
- Models:
    - Multiple Regression was explored to select the best predictive model.

Tools

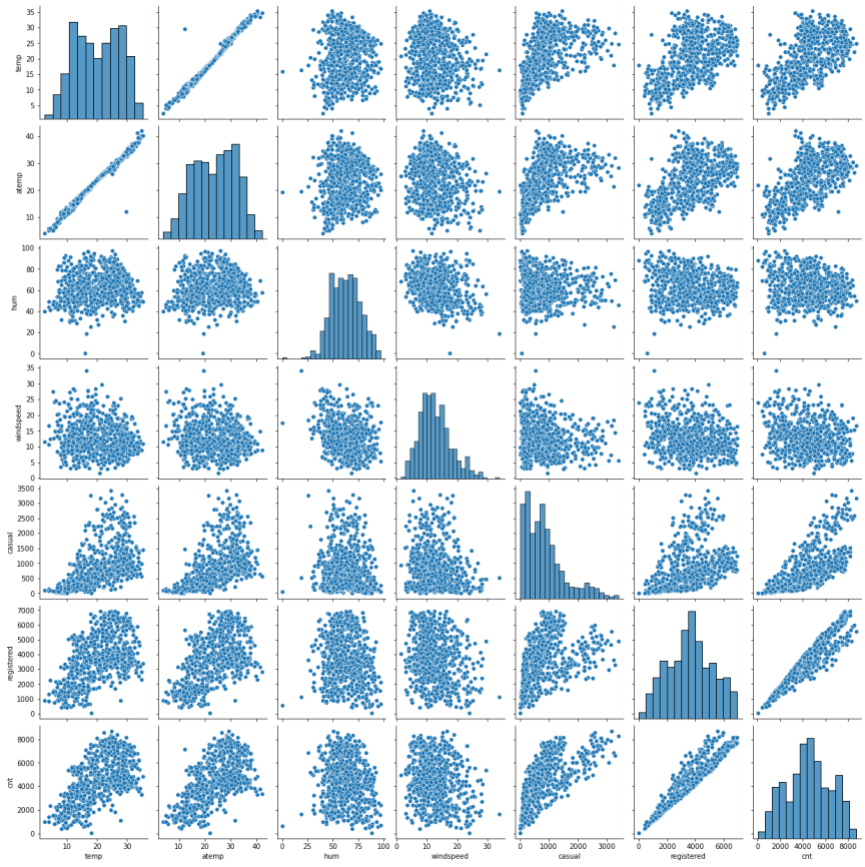
- Google Spreadsheets: Exploring, aggregating and summarizing the dataset
- Tableau: Visualizing the dataset to find helpful insights
- Python libraries:
    - Pandas and Numpy: Manipulating data
    - Statsmodels and Scikit-learn: Linear Regression

Communication

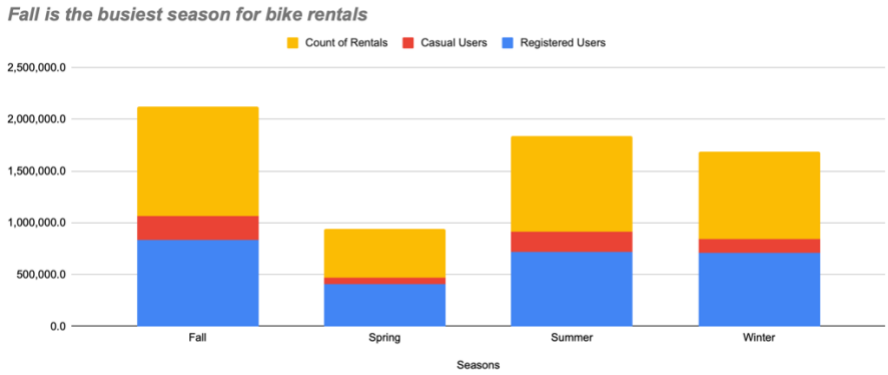
The final equation for the best fitted line for the regression model is:

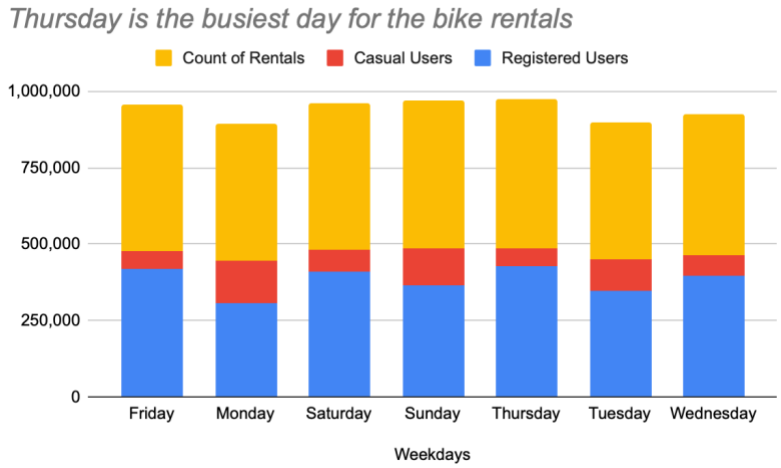
- cnt = 0.143 + 0.225 x yr -0.076 x holiday + **0.523 x temp - 0.1 x windspeed** + 0.084 x season_summer + 0.125 * season_winter - 0.072 x weather_cloudy - 0.261 * weather_lightrain + 0.092 * mnth_September
- This clearly indicates that temperature is the most important factor to predict the count of rentals.
- Windspeed has a negative correlation with the count of rentals.

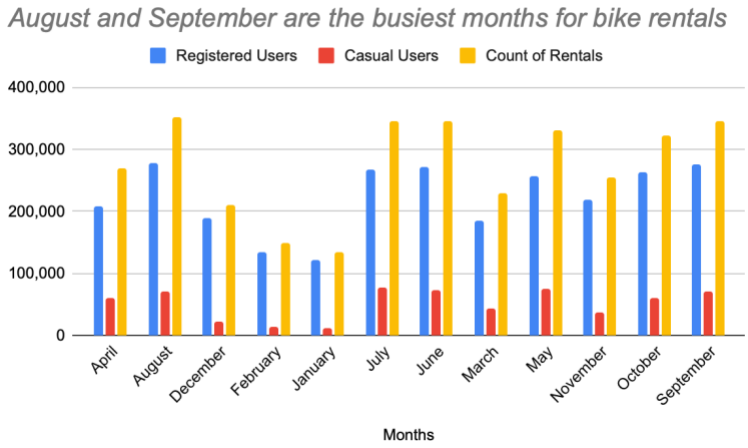
Pairplot of features and target variable

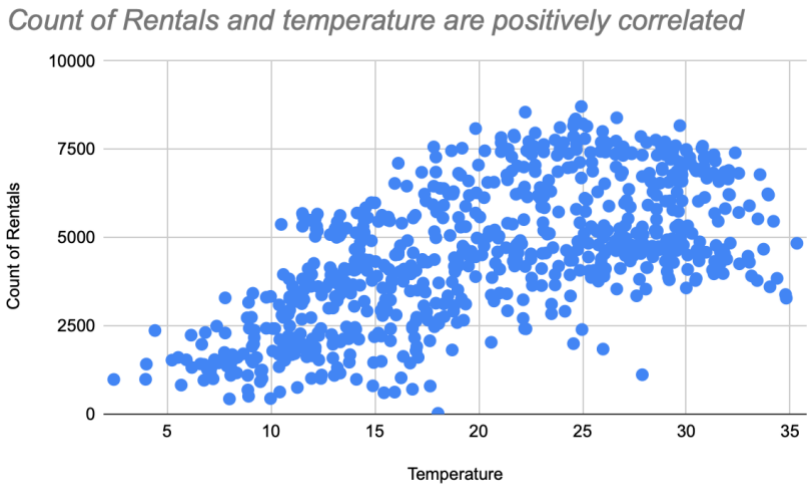


Following are the insights from Google Spreadsheets and Tableau:

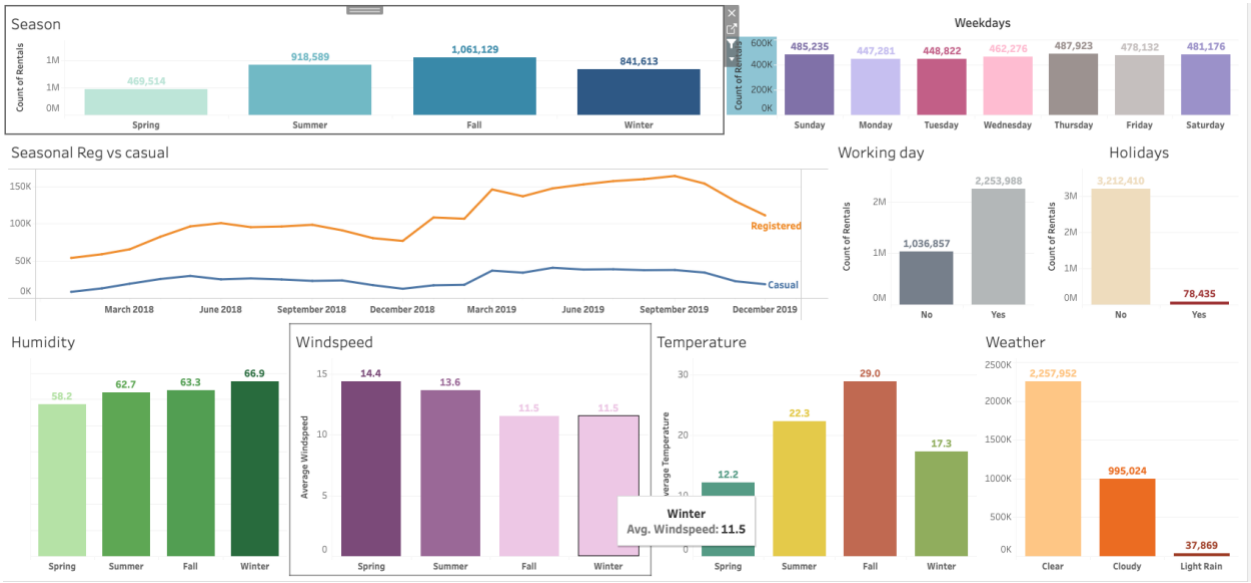




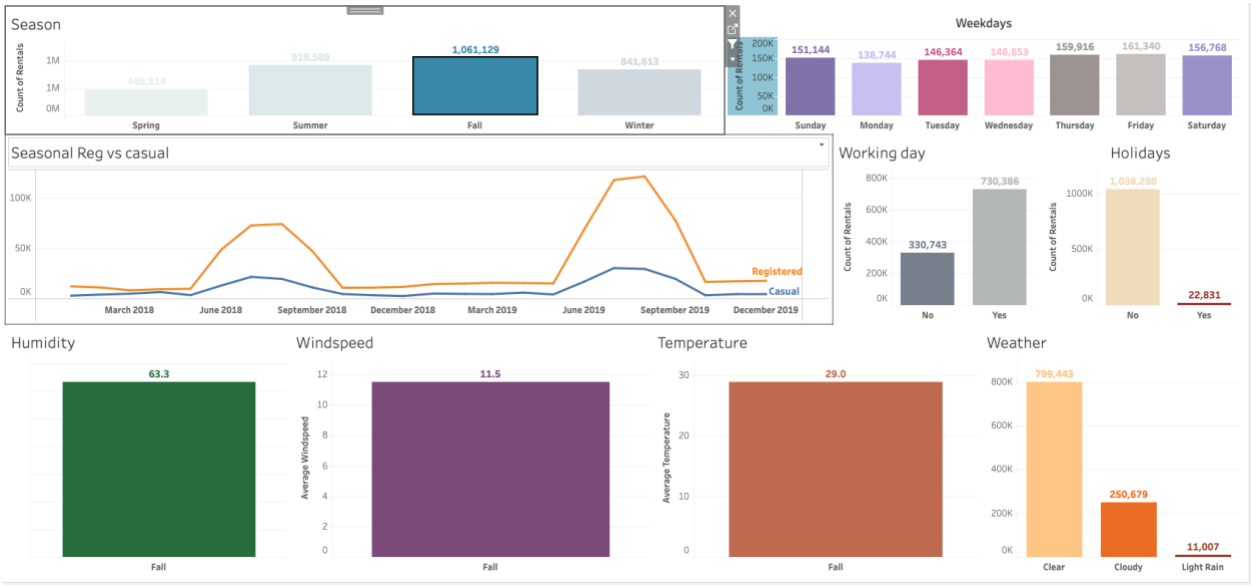


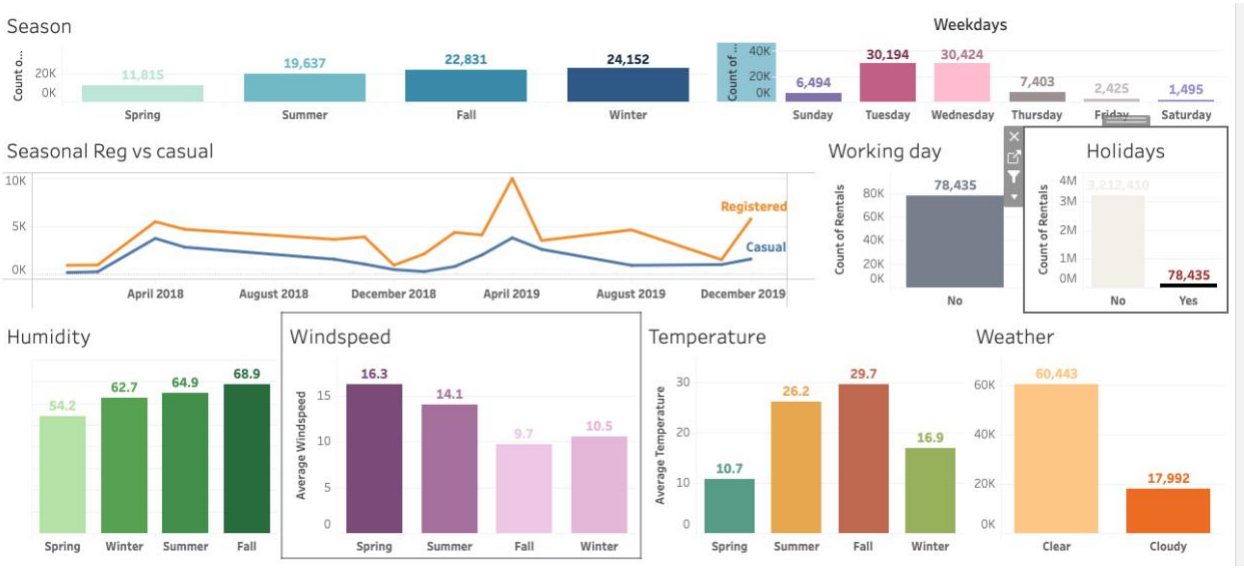


Interactive Dashboard



Fall Season is the Busiest Season



Registered Users Insights

Bikes are only rented under clear and cloudy weather for holidays