WelcomeBike LLC

Daily Bike Rental Projection Report

## presented by

James Lule

1. Background/Introduction

The purpose of this project is to build a neural network that will predict the number of bike rentals for a given day and hour for the WelcomeBike Bike Rentals Washington DC branch. In this report, we also aim to answer important stakeholder questions pertaining to the model creation, effectiveness of the model, the reasoning behind the different parameters used, and the biggest effect of certain features on revenue.

1. Stakeholder Questions

## Network Layers and Hyperparameters

*Question: How many layers should we use in our network?*

We were able to figure out through trial and error that one input layer of 256 neurons, with 4 hidden layers of 128 neurons each, and one output layer of one neuron gave us the best results. Having less layers and neurons seemed to cause the model to increase the mistakes it makes during its predictions. Having more layers and neurons seemed to have a similar effect, by increasing the mean squared error of the predictions.

*Question: If the model is underperforming, what hyperparameters will maximize the model’s improvement?*

As we experimented with the different hyperparameters, we as a team have decided that the number of layers and neurons within those layers is the most important hyperparameter when trying to determine how to maximize the improvement of the model because too few or too many can have negative results on the model’s ability to learn.

## Feature Engineering

*Question: How should the temperature features be handled?*

We took two measures to handle the temperature data. The first step was to normalize the temperature, which is just dividing by a constant to change each temperature to a number between 1 and 0 to balance the impact temperature had on the neural network. The second step was to MinMaxScale the temperature, which is just taking the min and taking the max and scaling the temperatures with the max as 1 and the min as 0. We performed the MinMaxScale upon all the features in the dataset, which ensured all the data was between 1 and 0.

*Question: What other feature engineering should we use for our data?*

We completed feature engineering on our data. The first was to create a feature of total users in a given day by adding the total casual and total registered together. The second was to one hot encode the seasons into the new features: winter, spring, summer, and fall. This gave us a binary yes-no answer to whether a certain entry was in a certain season rather than providing numbers 1 through 4, which might confuse the neural network to put more emphasis on some entries above others, providing an inaccurate prediction. The third was the same as the second, in which we one hot encoded the weathers into the new features: clear, cloudy, rainy, and heavy weather. This gave us a binary yes-no answer to whether a certain entry had a certain weather type rather than providing numbers 1 through 4, which might confuse the neural net.

## Learning Rate

*Question: What are we going to do to find the most optimal learning rate?*

Optimizers define the learning rate for a neural network. We explored the optimizers Adam, RMSProp, adadelta. We ended up using the Adam optimizer to define the learning rate for our neural network because the Adam optimizer built upon the other optimizers within its class, turning their weaknesses into strengths. This made the Adam optimizer the best all-around option for training models. We left the learning rate parameter at its default, which was 0.001, because we honestly didn’t think to change it. We could try that in the future.

## Loss Function

*Question: How will we know if our model has strong predictive power?*

The loss function compares the target and predicted output values. This helps us to know how well our model performs when compared against the training data. We will want to find the best loss function in order to maximize our model’s predictive power.

*Question: What do we plan to use for our loss function?*

In finding the total rented bikes for a given day, we have determined that we are dealing with a regression problem. That is because the total rented bikes is a real value that will be the output the model will give after analyzing season, weather and other features. After researching, we have determined the Mean Squared Error (MSE) will be the best value to use here. MSE is the average of the squared differences between predicted and actual values. As such, the results will always be positive. In addition, if the predicted and actual values match, then it will be a perfect 0.0, meaning there is no error. The squaring of the MSE means that the bigger mistakes the model makes compared to the actual value, the bigger the MSE will be. We are working with a real number output, which makes the MSE the perfect loss function for our model.

## Predictive Risk Model

*Question: In an effort to build a Predictive Risk Model, what are our recommendations?*

The best answer for this question would be that we can use AI to predict the likelihood of damage based on user profile data as long as the users have provided that profile data and we have kept record of previous damages. In the case of car insurance, these indicators are used in order to determine insurance rates, and so it would follow that bike insurance would work the same. Factors such as name, sex, age, and address are all used in determining insurance premiums for cars.

For example, a teenage male living in Los Angeles, California will have higher insurance premiums than a middle aged woman living in rural Idaho. Also, if someone gets in an accident, that information will be tied to that person’s name, which can be tracked. However, a machine learning model cannot be trained based on names, as that can potentially lead to disastrous civil suits and is unethical. Names should only be used to track individual history so that people who get in accidents will end up paying higher insurance premiums.

1. Stakeholder Focus Areas

*The stakeholders are particularly interested in the following areas:*

## Neural Network Results

Chart, line chart

Description automatically generated

After setting up the hyperparameters, and training our model, we found we are able to predict with minimal error the amount of bikes that will be rented per day. In the chart above, we can see the results of the model when comparing the model’s predicted total rented bikes per day to the actual rented bikes per day. The times where both lines match up show that our model predicted the actual value.

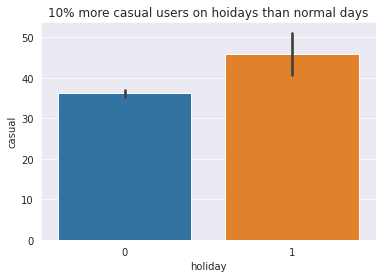
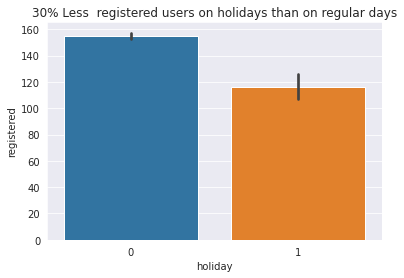
Based on this chart, we were able to find the R2 value, or the variance of the predicted values, with 100% explaining perfectly the variance of the model’s values. Our R2 score is 94.6%. While it is not perfect, our model is very good at predicting the data it has been trained to predict.

## Factors Affecting Revenue

*Question: Understand how weather, holidays, and other factors affect revenue.*

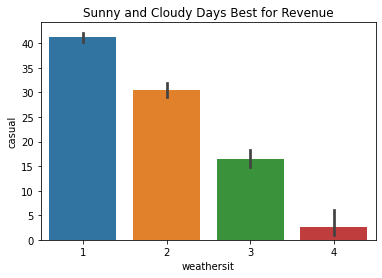
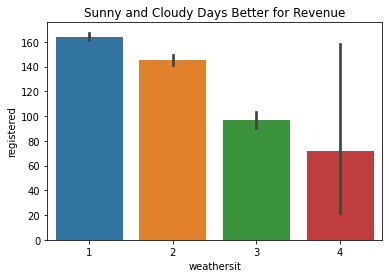
Since revenue data was not included in the data, we decided as a team to use the features for registered and casual users as an indicator of revenue. The logic is that when there are more users, revenue is higher, and when there are less users, revenue is lower. We found some interesting observations that we would like to share with the stakeholders.

**How Holidays affect Revenue: Registered Users vs. Casual Users**



It appears that there is a 25.3% increase in registered users on holidays compared to days that are holidays. As for casual users, even though there are less casual riders than registered riders, there is a 22.2% increase in riders on holidays compared to normal days. Assuming registered users use bikes to get to and from work, we think this means that registered users do not prefer to ride their bikes on holidays and that casual users prefer to ride their bikes on holidays.

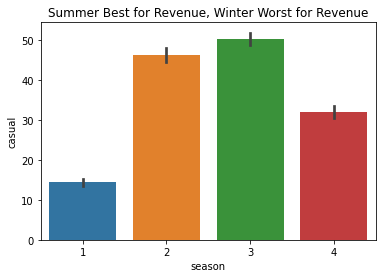
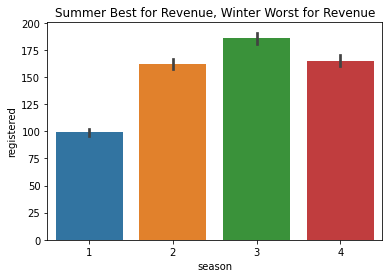
**How Weather affects Revenue: Registered Users vs. Casual Users**



*1: Sunny, 2: Cloudy, 3: Rainy, 4: Heavy*

There appear to be more people riding their bikes on sunny and cloudy days and less during rainy and heavy days. This trend is shown with the registered users and is even more pronounced with the casual users. We think this is because people prefer to ride their bikes when the weather is nice. This means revenue is better on days with better weather.

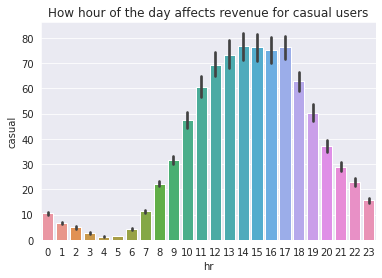
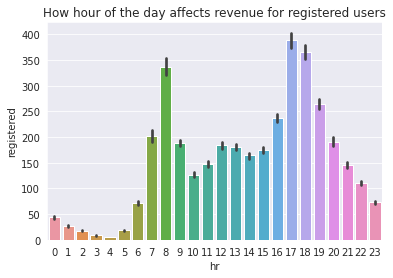
**How Seasons affect Revenue: Registered Users vs. Casual Users**



*1: Winter, 2: Spring, 3: Summer, 4: Fall*

There appear to be more people riding their bikes during the summer and less in the winter. This trend is more pronounced with casual users than with registered users. We think this is because people prefer to ride their bikes when the seasonal temperature is agreeable for riding a bike. This means that revenue is likely to be higher in the summer when the weather is warm and lesser in the winter when the weather is cold.

**How the Hour of the Day affects Revenue: Registered Users vs. Casual Users**



There appears to be different trends in the revenue between the registered and causal users, but the common factors between both which affect revenue are that most people like to ride their bikes when the sun is in the sky. However, registered users tend to drop between the hours of 9am and 4pm, while these are peak hours for casual users. This might be because registered users need the bikes to travel to and from work, whilst the casual users do not need the bikes during certain hours of the day. Thus, revenue is likely highest in the hours just before work starts for the day (7am-9am) and in the hours just after work (5pm-6pm) ends. And for the casual users there is gradual increase in riders with a sharp increase within the hours of (9am – 8pm) with the peak hours being (12pm – 6pm).

1. Python Notebooks

Below are Github Gist links to the notebooks we used during this case study:

James:

<https://colab.research.google.com/github/jlule/Machine_learning_module4/blob/main/Bike_rentals_Project_Module_4.ipynb#scrollTo=omrl_869JMrT>