The Models of London Bike Sharing Prediction

This manuscript (permalink) was automatically generated from Jingzi2020/CEE498_2020_G1@18a71fb on December 7, 2020.

Authors

- **Jingzi Chen**
  86-13522264917 · Jingzi2020
  Grainger College of Engineering, University of Illinois at Urbana-Champaign

- **Anye Wang**
  15409980880 · wanganye123
  Grainger College of Engineering, University of Illinois at Urbana-Champaign, Civil Engineering - Construction Management

- **Dana Monzer**
  0000-0002-4688-6359 · Dana2021
  Grainger College of Engineering, University of Illinois at Urbana-Champaign, Civil Engineering - Transportation Systems

Abstract

Bike sharing system has appeared more and more on the street of the cities in order to meet the demand of public transportation in the last short distance to the destination. Also, Bike sharing system is so popular around all the world that most of major modern cities and campuses have been operated. Among researches for the bike system, predicting the demanding number of future bike shares is one of the most important and necessary tasks. During our project, the aim of this study is to create a predictive model for bike-sharing counts in an hour in the city of London in United Kingdom. The model makes use of regular neural networks. And the main features affecting the bike counts include weather conditions and time variables. The model's root mean square is 210, with a mean of 1124 counts in an hour for testing data compared to 1138 counts in an hour in the training data. The model provides enough accuracy for planning the number of docks at a new station and scheduling bike redistribution schedules between stations.

Introduction

Bike-sharing system is an important mode in sustainable transportation modes. Great attention is put to improve such systems to increase their demand and maximize their environmental benefits along with other societal benefits. To study the demand of this system, London Bike-sharing system sis explored. The purpose of this study is to predict ranges of new bike counts each hour based on certain factors provided in the dataset.

The aim of this report is to create a predictive model using machine learning to predict bike counts in a given hour in London city which is useful for planning and operation forecasting of bike-sharing system. The exploratory data analysis using graphical and statistical tools in Python were used to derive preliminary conclusions about the dataset by analyzing the results of the tools used.

The dataset provided was acquired from three sources, to include the new bike counts in each hour, the weather conditions, and the holidays.

The data from cycling dataset is grouped by "start time", and it represents the count of new bike shares grouped by hour. The long duration shares are not taken in the count."

The data sample analyzed in this project is collected between January 1st, 2015 to January 1st, 2017 in London, UK and it includes the following parameters:

- Timestamp (year, month, day, hour)
- Cnt: the count of a new bike shares
- T1: temperature measure taken in degree Celsius
- T2: temperature feels
- Hum: humidity percentage
- Wind_speed: in Km/hr
- Weather_code: 1 = Clear ; mostly clear but have some values with haze/fog/patches of fog/ fog in vicinity 2 = scattered clouds / few clouds 3 = Broken clouds 4 = Cloudy 7 = Rain/ light Rain shower/ Light rain 10 = rain with thunderstorm 26 = snowfall 94 = Freezing Fog
- Is_holiday: 1 if it is a holiday, 0 if it is not.
- Is_weekend: 1 if it is a weekend, 0 if it is not.
- Season: 0: Spring, 1: Summer, 2: Fall, 3: Winter

The sections below include literature review, description of methods, results, discussion and conclusion.

Literature Review

Bike sharing system has appeared more and more on the street of the cities in order to meet the demand of public transportation in the last short distance to the destination. Also, Bike sharing system is so popular around all the world that most of major modern cities and campuses have been operated. Among researches for the bike system, predicting the demand of future bike shares is one of the most important and necessary tasks to ensure a satisfactory level of service of the system.

Many recent studies have focused on the problem of predicting demand of shared bikes in the system. In order to complete prediction, the bike sharing system is supposed to satisfy the real-time and high accuracy requirements. Some researchers utilized Spark Machine Learning framework to predict the number of rental bikes to optimize the accuracy of model [1]. Firstly, the author collected three types of data including Citi Bike, Weather and Holiday, which the data have been SQL and outlier processed. Then, three predicative models including multiple linear regression, decision tree and random forests are constructed to analyze and train the processed data. During the experiment, applying to machine learning Spark ecosystem, the author used 70 % data as training data while 30% as the test data. In the result, after testing, random forests model has the lowest root mean error (RMSE). At last, the researcher optimized the result further by applying logarithmic optimization to the model.

Throughout the literatures, there is consistent evidence that forecast number of shared bikes are influenced not by holiday, weather, and temperature [2] in general, but by user information such as gender, birth year and user type [1]. The strengths of the literature are that the author consider more factors like user information and process useful and reasonable row data. Moreover, the optimization improves the accuracy the result. While there has been much research on algorithm optimization, few researchers have taken the influence of different factors and the real-time data into consideration [2]. The weakness of the literature is that the author ignored the timeliness of data and station selection. Furthermore, the mythology is limited to predict the number in the new area where lack of the existing shared bike number. Usually, station clustering and demand prediction in every station should be evaluated and analyzed as an integration [3].

In general, it is challenging to collect the external factors in the future study because there are too many factors affecting users whether to use shared bike. However, we should consider multiple factors as much as possible to improve the accuracy. Comparing with our project's database, more factors are taken into consideration including time, number of new bike shares grouped by hour, temperature, humidity, wind speed, types of the weather, holiday, weekend, season [4]. In our project, we will employ historical usage of bikes with some necessary factors to predict the number, which are very similar with the literature because it is easier to collect data and the predicted numbers are closer to the real.

The second paper studied is "Case Studies on Transport Policy Modeling bike counts in a bike-sharing system considering the effect of weather conditions" [5]. This paper identifies a method to quantify the effect of weather conditions on bike sharing counts in San Francisco Bay area with the aim of improving bike sharing systems given their wide benefits such as decreasing transportation pollution and increasing mobility efficiency in cities, to name a few. The benefit of this model specifically will be to decrease the environmental costs and time consumption and other complications associated with the rebalancing operation of bikes between stations; which is important to ensure that each station has enough number of bikes to satisfy the demand, especially given the limited number of docks at station.

The methodology used to create the bike count model includes several steps. First the effect of various variables is quantified (month of the year, day of the week, time of the day and different

weather conditions), then these predictors were ranked by Random Forest technique and were used to predict a regression model using a guided forwarded step-wise regression. More than one model was created then the Bayesian information criterion was used to evaluate the models. In the first step, the count models employed generalized linear models, specifically two models were used Poisson, which condition to apply is that the mean and variance must be equal, and negative binomial regression, which uses same condition as Poisson except that another parameter is involved that loosens the initial condition and adjusts the variance independently, so it is considered to accommodate more dispersion. The second step after that is to apply machine learning to avoid overfitting of predictors, using Random Forrest method. The RF method randomly constructs a group of trees, where each tree is a subset of features, so trees are not correlated, then the ranking of features is obtained based on majority of votes from all trees, after that forward step-wise regression is applied, and finally a model is selected BIC after computing the log-likelihood of each model, and the model of the lowest BIC is to be selected.

This methodology [5] was applied on a dataset of bike-sharing for San Francisco Bay Area between August 2013 to August 2015, where incidents were documented every minute for 70 stations in the area, which led to a large dataset, and another dataset was used which included weather conditions during these 2 years, and it included the following attributes: “date (in month/day/year format), ZIP code, temperature, humidity, dew level, sea level pressure, visibility, wind speed and direction, precipitation, cloud cover, and weather description for that day (i.e., rainy, foggy or sunny).”

For the first step, the histogram of new counts frequency for all stations showed dispersion, which gave a hint on better fitness of NBRM. However, both PRM and NBRM were applied at first to generate a full model of all available predictors. Then RF was used to rank the predictors in the full model based on the OOB error. Forward stepwise regression was then used to fit several models that were constructed by RF, then BIC was applied to select the best subset of predictors to construct this model. At first it was assumed that there is no interaction between the 70 stations, to quantify effects fast and effectively and in an attempt to create one model for all variables rather than a model for each station, and this approach was described to satisfy the level of accuracy needed. And the results showed a logarithmic mean of bike counts at each station following parallel hyperplanes, which shows no interaction between stations. And in order to construct one model instead of 70 models, one for each station: 69 indicators were used with one reference, a similar approach was used for months of year with 11 indicators and January as a reference, and 6 indicators for the days of the week, and so on for all data attributes. If there was no significant difference between each pair of parameters, for example between 2 stations, it was assumed bike count was the same for the two stations to an acceptable level of accuracy. The results showed that different stations, month-of-the-year, day-of-the-week, and time-of-the-day were all shown to influence the model. And the following weather attributes were selected for additional exploration: mean temperature, mean humidity, mean visibility, mean wind speed, precipitation, and weather description. And for the second time, RF and forward step-wise regression were employed and the resulting models were compared by BIC. And the model with the trade-off between the minimum BIC value and the consideration of the effective parameters was selected. And as was shown because of dispersion, indeed NBRM was shown to be better than PRM, and it was selected for the rest of the modeling steps. Among 111 models created, the results showed that bike counts are significantly influenced by the month-of-the-year, day-of-the-week, time-of-the-day, and some weather variables, mainly temperature and humidity level, which is also dependent on geographic location. And the most significant variables affecting bike counts are available number of bikes at time t-1 and the time-of-the-day.

This paper holds many strengths in meeting our project, first the factors used in this paper match our dataset, and the paper justifies the use of these factors among others used in other studies. And this paper was the first to study the effect of humidity which was shown to have a significant effect on the model. The methodology and tools used match our set of expertise, which is to be developed through this course, and the tools and methods used were all explained and justified. This paper can be used

as a reference for us to build our model and compare the results given different factors, especially difference in geographic location, while also contributing to this research with the insights we obtain.

However, a few weaknesses of this paper were detected. First, the paper targets only docked bike-sharing model and it's applicable to only certain geographic areas with certain weather conditions, as for another location, different parameters might be additionally considered. The final stages of refining the model and detecting possible errors was merely systematic and rather relied on observations and experts opinions, so in addition to the scientific and sequenced steps, there was some subjectivity in the methodology when it comes to certain decisions like the number of trees in RF and the selected factors for each tree and their number which wasn't discussed and explained enough, so it might be hard for us to follow the same methodology at these stages. The paper only used Poisson and negative binomial models at the first stage and didn't attempt more complex distributions. And the paper chose the final model prioritizing simplicity, designated by a smaller number of predictors, among the last two proposed models, without explicitly comparing their levels of accuracy. And some standard values of comparisons for example minimum log-likelihood and typical BIC measures weren't shared in the paper.

Another study published by Lin, Wang, Jiang, Fan and Sun (2017) [2], tried to figure out sharing bike demand's prediction based on the Bayesian classifier and APSO-BP neural network models. In this article, the researchers collected weather data and historic public bicycle sharing record to build the dataset. Then, the model categorize bicycle rental mode through Bayesian classifier and used the specific neural network fitted to the mode. The evaluation of this model was based on its accuracy and the result was that the model showed higher accuracy than other algorithms. To find the valuable factors of bike rental prediction, the article discussed the influence of holiday, weather and temperature and decided to build the dataset with these significant factors. The article set "di" as 1, 2 and 3 representing weekday, the first half of holiday and the second half of holiday, "wi" as 1, 2 and 3 as different weather and "ti" as different ranges of temperature. To classify different situation of bike rental, the authors used cluster analysis to analyze the modes of bike rental record in training dataset. Through K-means cluster analysis, the bike rental behaviors were divided into 4 modes. Then, the researchers built the classifier using training dataset through Bayes classification method. Then, the article designed an APSO-BP neural network model to forecast the bike rental demand. APSO-BP used single hidden layer neural network model. APSO algorithm generated several particles which are feasible answers randomly at first. The algorithm used the squared error to evaluate fitness. The researchers trained separated ASPO-BP neural network for each mode. The experiment of this research was based on the bike rental record and the weather data in Hangzhou from March 18 to June 15 in 2016. 5 minutes were chosen as the time range. The researchers categorize the training dataset into 4 modes through K-means cluster analysis and then trained ASPO-BP neural network for each mode. The result showed that the accuracy of the demand predicted by this method was influenced by the accuracy of classification and the accuracy of this method in predicting rental demand was higher than other common methods. From this research, I learned that the factors should be transferred into reasonable values firstly and the method of transferring the data are diverse. Besides, the simple neural network models sometimes cannot solve the problem accurately. The combination of different algorithm are needed. Moreover, there are many optimized methods thus researchers should choose the most suitable optimized method to improve the accuracy. These tips can be useful for our project.

Exploratory Data Analysis

Introduction

Exploratory Data Analysis is vital and necessary before we start to create a training model because it helps us to realize and evaluate the both the features of data and their correlation with each other.

There are five main tools were used to describe the dataset.

Scatterplots

The reason for using scatter plots is to observe relationships between each pair of variables. Each dot in a scatter plot reports the values of individual data point, all the datapoints when plotted can help us identify a pattern that can show whether a relationship can be derived between two variables. This tool will help us decide which pairs of variables to explore further through other tools. Each pair of variables will have two plots.

Boxplots

The reason for using box plots is to provide a visual summary of the data enabling us to visualize the median and range values, the dispersion of the data set, and any signs of skewness. So we can compare dispersion of new bike counts with respect to different values of a parameter. This tool will be used to explore certain relationships further than scatterplots, and not for all pairs of parameters. A box plot includes five values: the minimum value, the 25th percentile (Q1), the median, the 75th percentile (Q3), and the maximum value.

Multivariate Point Plots

The reason multivariate point plots are used is to shows us the effect of certain variables on the distribution of the new bike counts during the day. So we can obtain different plots depending on the other variable, so we can explore two variables interrelations. And we choose categorical variables to find how their different categories affect the change in new bike counts along a day.

Statistical Values

The reason for finding statistical values is get an insight about the mean, range and maximum and minimum values of different parameters especially new bike counts, and make a connection with the visualizations created. If the dataset included more than 2 years, a good approach would have been to also find covariance between datapoints in different years to identify any change in pattern resulting from certain events associated with a certain year, but this is not needed in our case.

Correlation Matrix

The reason for using correlation matrix is to identify correlation between all parameters in the dataset and the ones affecting new bike counts only. This is important for the predictive model, so we understand and interdependencies between predictors and to avoid multicollinearity. Correlation values range between -1 and 1., positive values indicate positive correlation and negative values indicate negative correlation. Absolute Values above 0.7 generally indicate high positive correlation, between 0.5 and 0.7 is moderate and between 0.3 and 0.5 is low, while under 0.3 is negligible correlation.

Data Visualization

Scatterplots

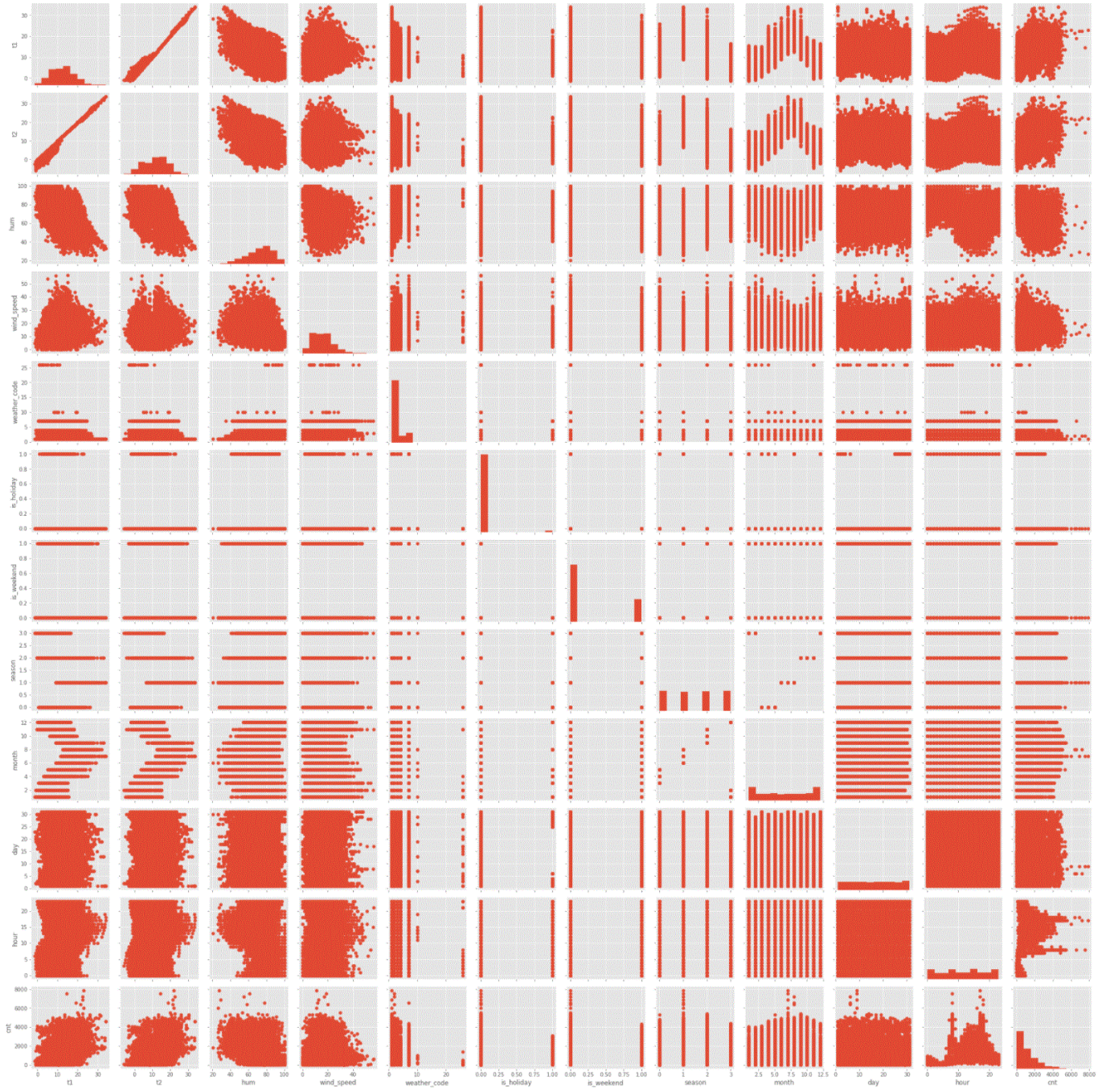

Figure 1: Scatterplots

By looking at the plots we observe the following:

- 1- t1 and t2 are linearily dependent, which is expected because temperature feel directly depends on temperature.
- 2- Wind_speed and t2 seem to have some kind of relationship, which needs to be explored further
- 3- t1 and t2 have some kind of relationship with humidity, which needs to be explored further.
- 4- t1, t2, humidity and wind speed have some kind of relationship with cnt, which needs to be - explored further.
- 5 - Weather code and cnt have some kind of relationship.
- 6- Cnt seems higher when is_holiday is 0 (so when it is not a holiday)
- 7- Cnt seems higher when is_weekend is 0 (so during weekdays)
- 8- Cnt seems higher for weather codes (0-9)
- 9- Cnt seems higher for season 1 (summer), then similar for seasons 0 and 2 and lowest for season 3 (winter).
- 10- Cnt seems highest for month 7,9,8,5 (expected based on observation 9)

- 11- It doesn't seem a certain pattern can be derived for effect of day on cnt.
- 12- Cnt seems to have a relation depending on hour that can be explored further and it seems to be a multi-peak distribution

The observations above allow us to build hypthesis that can be explored further using other EDA tools which will lay down the basis of our predictive model.

Boxplots

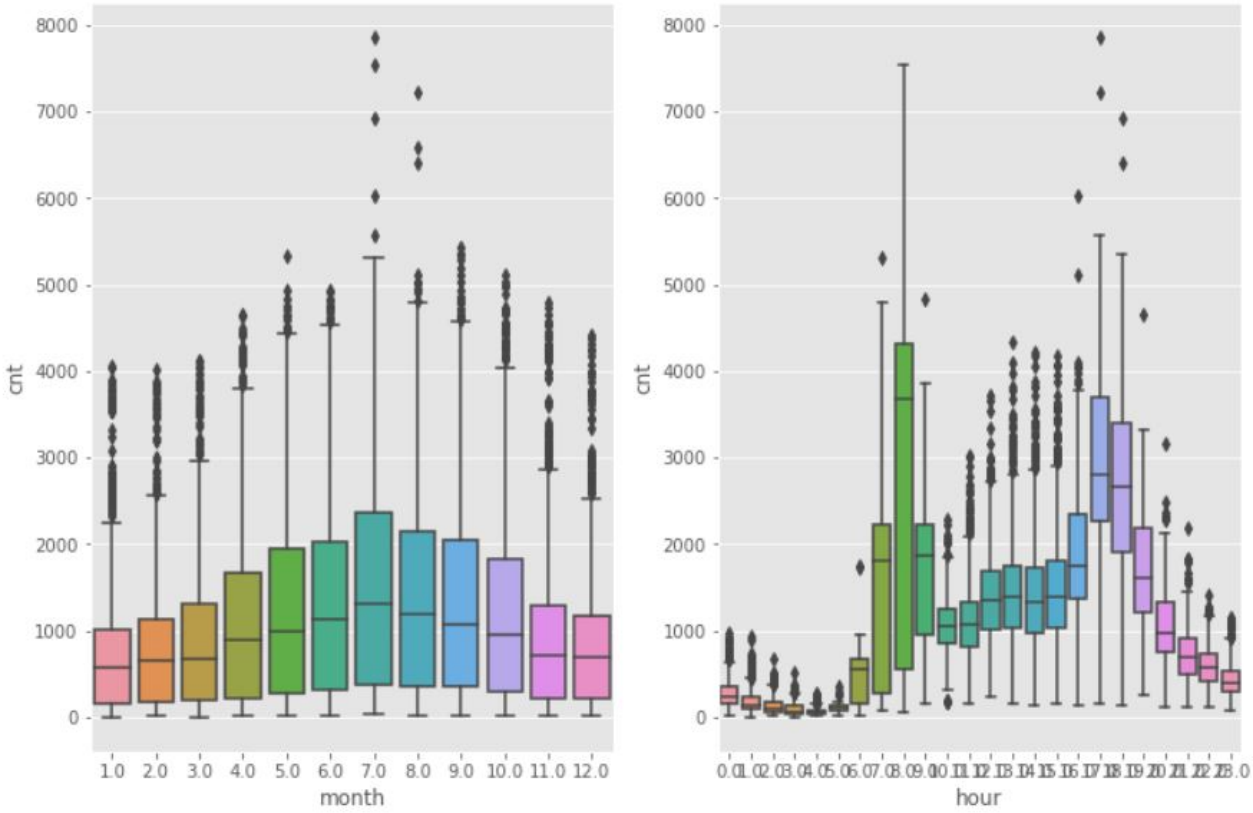

Figure 2: Boxplots of Data

Graph1 in Figure 2 shows that:

- 1- There seems to be a distribution of the median (middle line of the boxplot) and range (of the black line of each boxplot) of cnt with respect to month, it roughly follows a normal ditribution, peaking in July.
- 2- For each month the distributions tends to be skewed to the right.
- 3- Months with lowest ranges are January, February, November and December.

Graph2 in Figure 2 shows that:

- 1- There is no clear distribution of median of cnt over the hours of the day, but the variation shows that cnt varies between different hours in the day, which seems to have two peaks, around 8 am then around 5 pm.
- 2- During some hours, the shape of the distribution of cnt tends to be normal, while it's more skewed to the right for some other hours.
- 3- As expected night hours have the lowest cnt.

Multi-variate Plots

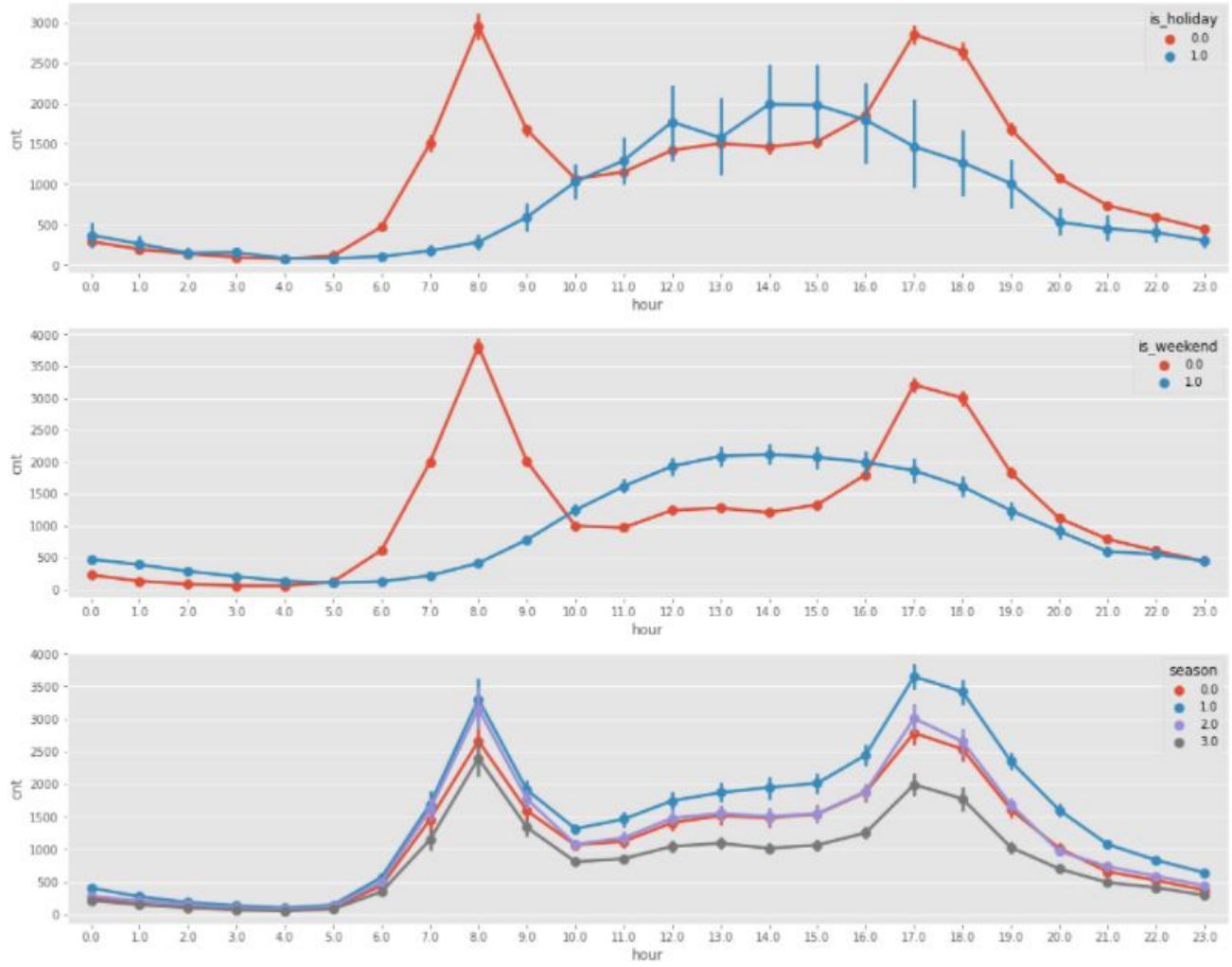

Figure 3: Multi-variate Plots of Data

Figure 3 shows the plots of average bike counts at each hour of the day. The following trends are shown:

- The distribution of the counts during the day is different on a holiday is at peaks around 1 pm - 2pm, while during regular days, it has two peaks around 8 am and 5 pm, which follows the governing distribution and was shown in the boxplots as well.
- There are similar trends for weekends as holidays and for weekdays as non-holiday days
- The seasons order based on highest bike counts is summer, then fall and spring, and then winter, which validates more clearly previous results.

Statistical Values

	cnt	t1	t2	hum	wind_speed	weather_code	is_holiday	is_weekend	season	month	day	hour
count	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000	12223.000000
mean	1138.098830	12.481742	11.543715	72.407081	15.923757	2.730426	0.020944	0.282091	1.498405	6.526139	15.739344	11.464289
std	1079.326816	5.546247	6.588240	14.262836	7.925239	2.295948	0.143203	0.450036	1.120819	3.462204	8.802342	6.903442
min	0.000000	-1.500000	-6.000000	20.500000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
25%	255.500000	8.500000	6.000000	63.000000	10.000000	1.000000	0.000000	0.000000	0.000000	4.000000	8.000000	6.000000
50%	841.000000	12.500000	12.500000	75.000000	15.000000	2.000000	0.000000	0.000000	2.000000	7.000000	16.000000	11.000000
75%	1670.500000	16.000000	16.000000	83.000000	20.500000	3.000000	0.000000	1.000000	3.000000	10.000000	23.000000	17.000000
max	7860.000000	34.000000	34.000000	100.000000	56.500000	26.000000	1.000000	1.000000	3.000000	12.000000	31.000000	23.000000

Figure 4: The Statistical Values of Data

Figure 4 gives us general insight about the mean and range of each parameter but it's irrelevant for time parameters. The main outcomes is that the values of cnt in each hour over the two years of

study, have a mean of 1138 and standard deviation of 1079, which is almost equivalent to the mean which means that there is significant difference between cnt values in certain hours than others. And this was indeed shown in the visualizations. We also notice that the maximum value is 7860, almost 7 times as big as the mean, which also means that there are very relatively few hours with such high values that they didn't affect the mean significantly. However this might be also affected by the fact that we the datapoints include records along the day including night time, were people are not using bikes or commuting at all. We notice the temperature values range between -1.5 and 34 degrees C, which might not be representative of others cities; same applies to teh humidity and wind speed. So it's important to highlight this factor when representing the predictive model at later stage.

Correlation Matrix

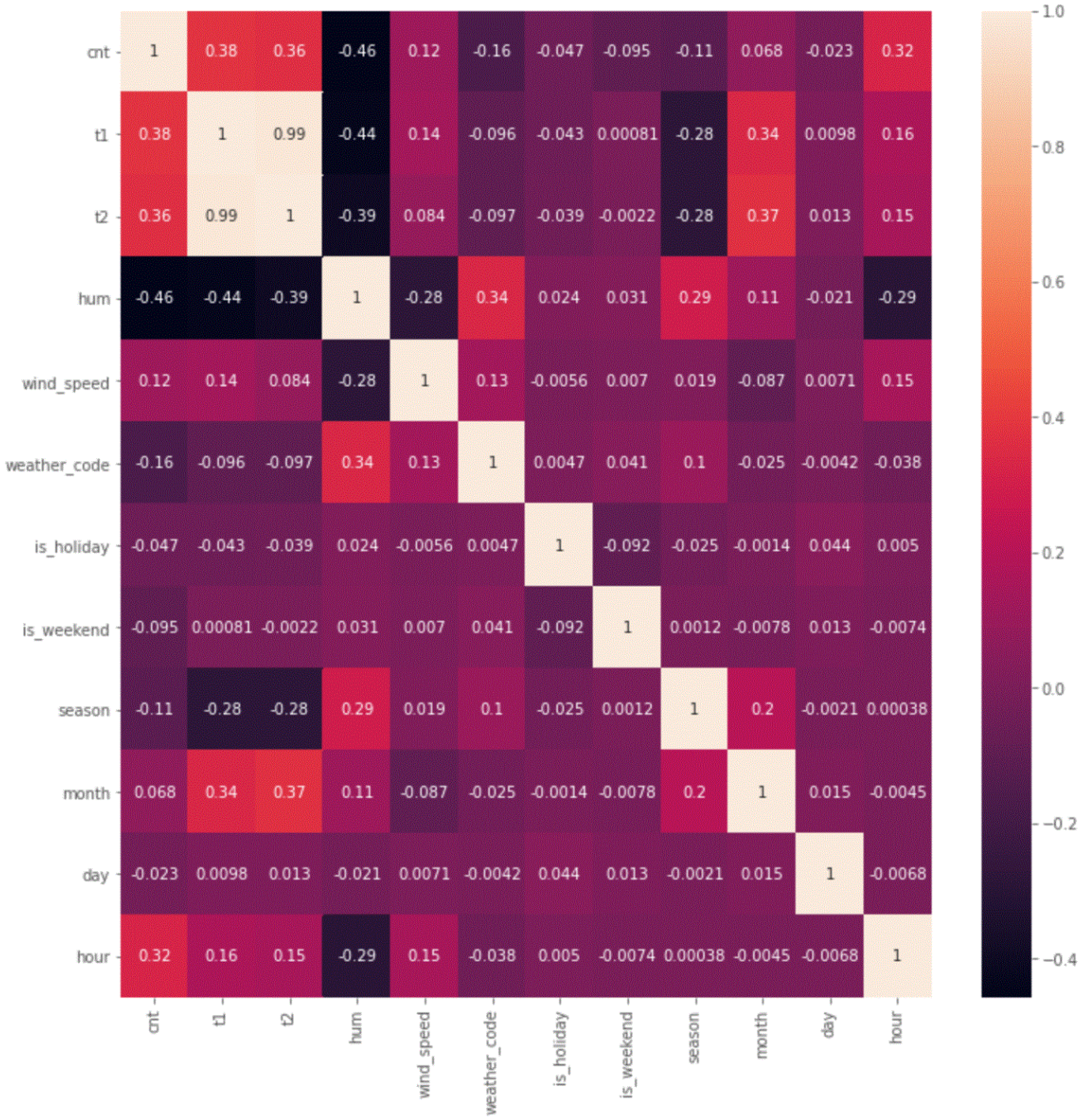

Figure 5: The Correlation Matrix of Data

We can follow the color map 5 to find the variables of some correlation, which are the variables with correlation above roughly above 0.3, which are colored by red or lighter shade of color or less than -0.3, which are colored in black . As expected t1 and t2 have high correlation which means only variable might be included in the predictive model at most.

- cnt shows correlation to t1 & t2 & humidity and hour of the day.
- t1 and t2 show correlation to cnt, humidity, and month of the year.
- humidity shows correlation to cnt, t1 & t2, and weather code
- wind speed shows correlation to slightly to humidity
- weather code shows correlation to humidity
- holidays binary doesn't show correlation to any other parameter
- weekend binary doesn't show correlation to any other parameter
- season shows correlation to t1 & t2, humidity
- month shows correlation to t1 & t2
- day of the month doesn't show correlation to any other parameter
- hour shows correlation to cnt and humidity

In a word, the scatterplots showed that t1 and t2 are directly related, and possible relationship between wind speed and t2, t1 and humidity, humidity and bike counts, bike counts and wind speed, bike counts and temperature, weather code and bike counts. It also shows that bike counts are higher during weekdays and non holidays and for better weather conditions. Also, it shows that highest counts occur for summer. And no pattern was shown between counts and day of month. The boxplots showed that there seems to be a distribution of counts based on month, and that the counts are mostly skewed to the right. Also, that the distribution of counts over the hours of the day have two peaks, and within the hour the distribution is mostly normal but for some hours it is skewed to the right. Multi-variable point plots show interesting insights about the change of the variation of counts over the day from bi-modal, for regular days and weekdays, to unimodal for holidays and weekends. And it also showed that although summer has highest counts and winter has the lowest, fall and spring seem to have similar values. And the correlations showed that new counts are mainly correlated to temperature, humidity and hour of the day with no significance correlation to other parameters. And some of the other parameters are correlated as well especially weather attributes (humidity, temperature, wind speed..).

We can derive several conclusions from the results of the exploratory data analysis. We explored which factors have a direct relationship with the new bike counts registered each hour, and the results show that the main factors affecting the range of new bike counts in an hour are temperature, humidity, and hour of the day, with some effect shown in visualizations by the weekdays, weather description, month and season.

The Models to Predict the Amount of Bike Sharing

Regular Neural Network

Based on the exploratory data analysis, regular neural network is used to figure out the project.

Neural network is a model that optimize the parameters through learning process to recognize hidden relationships between different data.

The algorithm used for this problem is regular neural networks. Because of the low correlation between given features, neural network probably is the most appropriate model to solve the project. This model is used for similar predictions to predict a dependent variables based on independent variables, where the model is unlikely to be a linear regression model, unlike convolutional neural

networks that are used for image classification, or recurrent neural networks that are used for prediction of future outcomes that are dependent on past data, like predicting the left life span of equipment, which is irrelevant for this problem as the new bike counts are independent of past bike counts, unless available bikes in a station in the previous hour was provided as a feature as literature have shown it can affect bike counts, however this information is not available in the dataset. However, another algorithm that is applicable for this problem is Random Forest Model, however this algorithm wasn't used either. The reason regular neural network was preferred is because random forest has a few limitations. The main limitation of random forest is that a large number of trees is required to produce accurate predictions, but on the other hand, the higher the number of trees used in the model the slower it is [6]. Moreover, in reference to Ashqar et al. (2019)[5], feature engineering in random forest becomes even more critical which adds another difficulty in improving accuracy of the model besides hyperparameters tuning.The architecture of regular neural network is shown in Figure 6 .

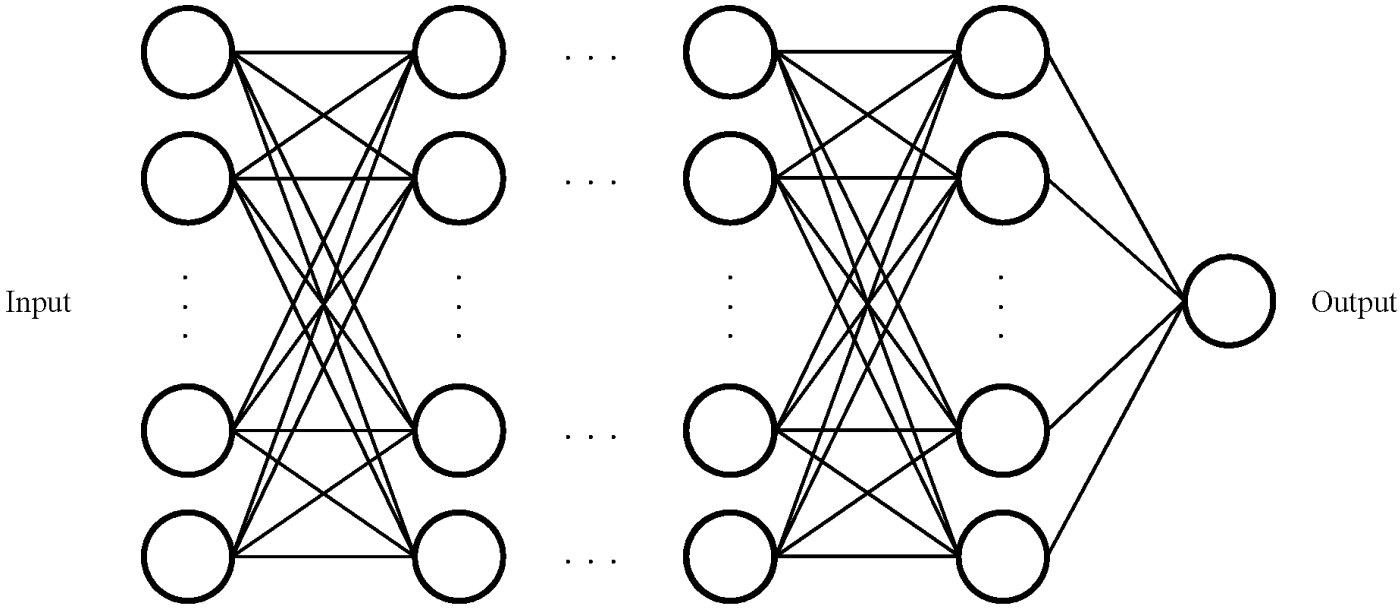

Figure 6: The Architecture of Regular Neural Network

There are 4 steps to build and train neural network, including:

- Selecting features;
- Data preprocessing;
- Designing layers and parameters;
- Determining training methods.

We used different features, epochs, hidden layers, units and learning rates in this project. The architectures of our models are shown in Figure 7 , 8 and 9 . The main parameters of our models are shown in Table 1

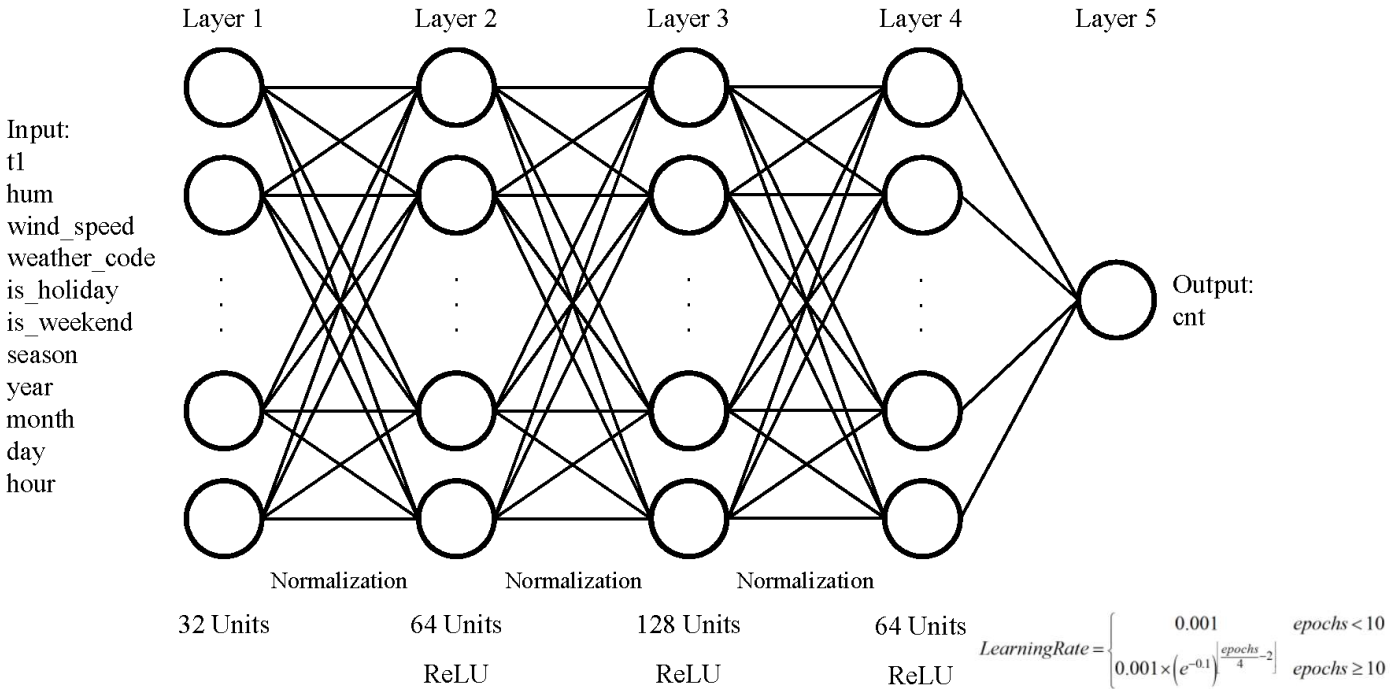

Figure 7: The Architecture of Jingzi's Model

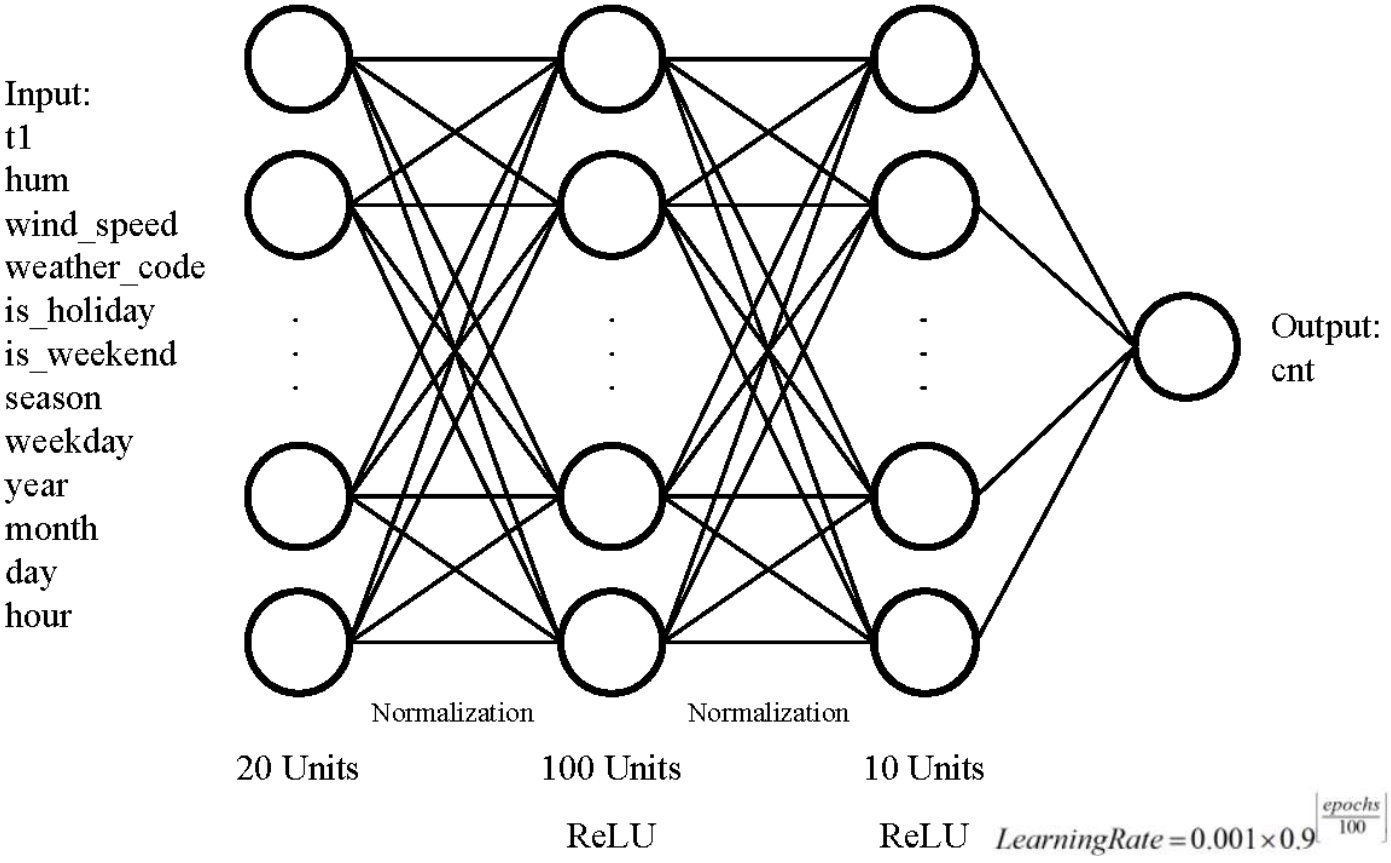

Figure 8: The Architecture of Anye's Model

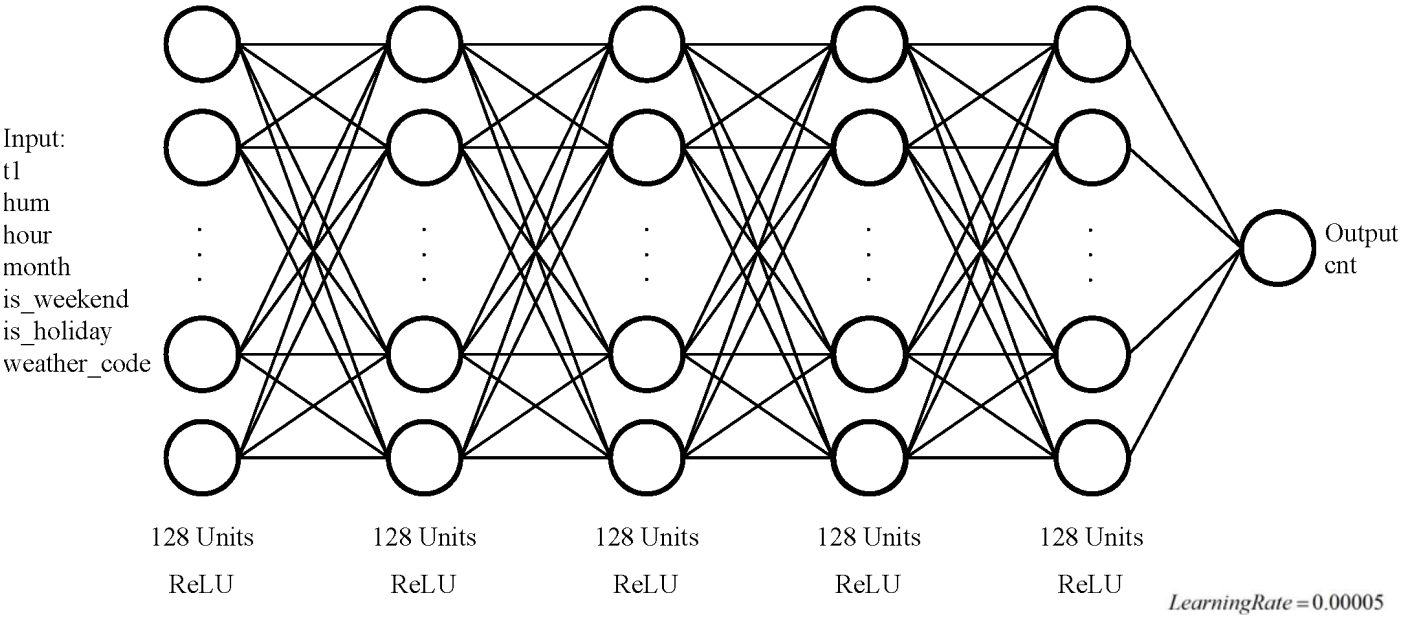

Figure 9: The Architecture of Dana's Model

Table 1: The Overview of Our Models

Models	Features	Epochs	Hidden Layers	Units	Learning Rates
Jingzi	11	<150	5	289	Non-Constant
Anye	12	300	4	131	Non-Constant
Dana	7	800	6	641	Constant
*Epochs in Jingzi's model will be stopped when loss reached the minimum.					

ReLU is used as activation function because ReLU is not only usually better than other activations like the sigmoid, but also easier to compute [7] . The equation of ReLU is shown in equation 1

$$F(x) = max(0, x) \tag{1}$$

The loss function and evaluation of the models' performances are based on the root mean squared error(RMSE) between the test data and predictions. The equation of RMSE is shown in equation 2 .

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(observed_i - predicted_i)^2} \tag{2}$$

In equation 2 , "n" is the number of predictions. "observed_i" and "predicted_i" is the observed cnt and the predicted cnt at each group of features. Thus, RMSE represents the differences between observations and predictions.

Sensitivity Analysis

To evaluate and optimize our models, we analyzed the sensitivity of different parameters.

Units of Layers, Layers, Normalization and Learning Rates are analyzed. To avoid the influence of randomization, the evaluation of the performances is according to the average RMSE of 3 separate

training with same initial parameters.

The architectures and main training methods of sensitivity analysis are shown in Table 2 , Table 3 , Table 4 and Table 5 .

Table 2: The Sensitivity of the Number of Units in Layer 2

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization
32	16	128	64	1	Same with Jingzi	At each layer
32	32	128	64	1	Same with Jingzi	At each layer
32	64	128	64	1	Same with Jingzi	At each layer
32	128	128	64	1	Same with Jingzi	At each layer

Table 3: The Sensitivity of the Number of Layers

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization
32	64	128	64	1	Same with Jingzi	At each layer
-	64	128	64	1	Same with Jingzi	At each layer
-	-	128	64	1	Same with Jingzi	At each layer

Table 4: The Sensitivity of Normalization

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization
32	64	128	64	1	Same with Jingzi	At each layer
32	64	128	64	1	Same with Jingzi	-

Table 5: The Sensitivity of Learning Rate

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization
32	64	128	64	1	Same with Jingzi	At each layer
32	64	128	64	1	Same with Anye	At each layer
32	64	128	64	1	0.001	At each layer
32	64	128	64	1	0.005	At each layer

Results of Modeling

Performances of Different Approaches

Using the above methods, we get the following results.

RMSE represents the differences between the predictions and real data. Thus, smaller RMSE indicates that the performance of corresponding model is better. The results in Table 6 indicate that Jingzi's model has the best performance in prediction.

Table 6: The Average RMSE of Each Models

Models	Features	Epochs	Hidden Layers	Units	Learning Rates	RMSE

Models	Features	Epochs	Hidden Layers	Units	Learning Rates	RMSE
Jingzi	11	<150*	5	289	Non-Constant	227.702
Anye	12	300	4	131	Non-Constant	314.257
Dana	7	800	6	641	Constant	248.733
*Epochs in Jingzi's model will be stopped when loss reached the minimum.						

Thus, the sensitivity analysis are based on Jingzi's model because of its performance in RMSE.

Performances of Models with Different Number of Units in Layer 2

In the sensitivity analysis, the number of units in Layer 2 are changed. The performance of models with different number of units in Layer 2 are shown in Table 7 .

Table 7: The Performance of Models with Different Number of Units in Layer 2

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization	Average RMSE
32	16	128	64	1	Same with Jingzi	At each layer	246.594
32	32	128	64	1	Same with Jingzi	At each layer	241.033
32	64	128	64	1	Same with Jingzi	At each layer	232.210
32	128	128	64	1	Same with Jingzi	At each layer	243.432

Comparing models with different units, the model which has 64 units in Layer 2 had the lowest average RMSE. The results indicate that setting more units in Layer 2 can improve the performance of model in this test dataset. However, the RMSE increase when units in Layer 2 are more than 64.

Performances of Models with Different Number of Layers

Then, the number of layers are changed to analyze the influence of layers. The performance of models with different number of units in Layer 2 are shown in Table 8 .

Table 8: Performances of Models with Different Number of Layers

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization	Average RMSE
32	64	128	64	1	Same with Jingzi	At each layer	232.210
-	64	128	64	1	Same with Jingzi	At each layer	313.591
-	-	128	64	1	Same with Jingzi	At each layer	693.552

Comparing models with different number of layers, the model which has 5 layers had the smallest average RMSE. The results indicate that adding more layers properly can improve the performance of model in this test dataset.

Performances of Models with or without Normalization

Besides, the influence of normalization are analyzed. Normalization in Jingzi's model was removed. The performance of models with and without normalization are shown in Table 9 .

Table 9: Performances of Models with or without Normalization

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization	Average RMSE	
32	64	128	64	1	Same with Jingzi	At each layer	232.210	
32	64	128	64	1	Same with Jingzi	-	251.478	

Comparing models with or without normalization, the model which include normalization perform better. The results indicate that normalization can improve the performance of model in this test dataset.

Performances of Models with Different Learning Rates

Last but not least, learning rates of Jingzi's training method are changed. The performance of models with different learning rates are shown in Table 10 .

Table 10: Performances of Models with Different Learning Rates

Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Learning Rate	Normalization	Average RMSE
32	64	128	64	1	Same with Jingzi	At each layer	232.210
32	64	128	64	1	Same with Anye	At each layer	222.960
32	64	128	64	1	0.001	At each layer	223.531
32	64	128	64	1	0.005	At each layer	231.112

Comparing models trained in different learning rates, the methods setting 0.001 as learning rate and setting learning rate scheduler of Anye perform better in RMSE. The methods setting 0.005 as learning rate and setting learning rate scheduler of Jingzi require less epochs. The results indicate that learning rates can influence the performance of model in this test dataset. Moreover, non-constant learning rates or bigger learning rates can shorten the training time in this training dataset. To optimize the initial model, the learning rate scheduler can be changed. The improved smallest RMSE is 210.657 and its average RMSE is 222.960.

Discussion

The three models obtained by the group members were compared and assessed. And the best-fitting model was used to predict bike counts in an hour for the test data and it resulted in the lowest denormalized RMSE equivalent to 210, however the other models had close RMSE as well. The selected model includes the following features: temperature, humidity, hour of the day, day of the month, month of the year, the year, weekend indicator, holiday indicator, weather code, season, & wind speed. To asses if the RMSE value is acceptable, we look at the scale of the bike-sharing system in London. Transport for London website shows that there are more than 11,500 bikes at over 750 docking stations across London. So 210 bikes error in estimation for all stations is equivalent to an overestimation of underestimation of bike rentals by 0.28 per station, which corresponds to 28%

chance than 1 bike more or less will be used per station, assuming demand over stations is uniform since the dataset didn't include information about where the bike counts where recorded in the city and in which station. In the paper by Ashqar et al. (2019) [5]. the mean prediction error (MPE) which is equivalent to the sum of the difference between predictions and actual values divided by total number of datapoints, varies between -1 and 2.2 per station, and in this case the MPE value we obtained is 0.28 per station, which means that the model has high prediction power. The purpose of this model is to predict the future bike share of London bike-sharing system. Since the demand is stochastic, regardless of how well the model performs there is always a level of uncertainty, so a range of the bike new counts is the objective for this problem. We observe that the mean of the training data is 1138 and the testing data led to a mean of 1124, which shows a good level of accuracy, with only 1.23% underestimation of the mean of bike counts. In another model the mean of the bike counts for the testing data was equivalent to 1167 which is equivalent to 2.54% overestimation of the mean. In some cases, overestimating the demand is preferred rather than underestimating the demand, thus both models can yield satisfactory predictions. Besides that the coefficient of determination (R-squared) has a value of 0.975 which is another validation that the model's performance is satisfactory. The model can be used to predict number of bikes in a station per hour, which can be used to schedule bike rebalancing between stations or to design a new station and estimate the number of bikes required, however for this task additional dataset that includes distribution of the demand over stations is required. The level of accuracy achieved in this model, is appropriate to set shifts scheduling such that demand is met by ensuring a certain number of bikes is available at each station at a certain period during the day. Given the safety margin added in such operations to account for stochasticity of the demand, the model performance can be extremely efficient for such applications. However, it is not ideal to use this model for scheduling tight shifts in busy areas where the demand can shift drastically and the bike count per hour is the highest and more accuracy is needed given that this model assumes uniform demand over all stations. The actual bike counts of the testing data are not provided, so to compare the actual bike counts with the predicted bike counts, we use the validation data which constitutes about 30-40% of the training data chosen at random, and we predict the corresponding bike counts. Two plots are created using the actual bike and predicted bike counts with respect to the month Figure 10, which shows that the average actual bike counts and predicted bike counts don't vary significantly with respect to each month, same applies for the second plot Figure 11 that plots the counts per timestamp which includes all the datapoints for all the hours in the validation data, and similarly we can see that the predictions have high accuracy.

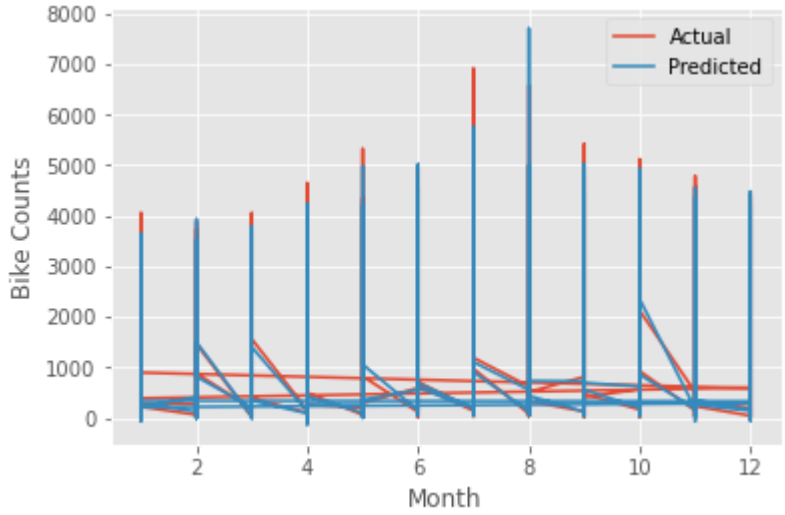

Figure 10: Plot of the actual and predicted bike counts with respect to month

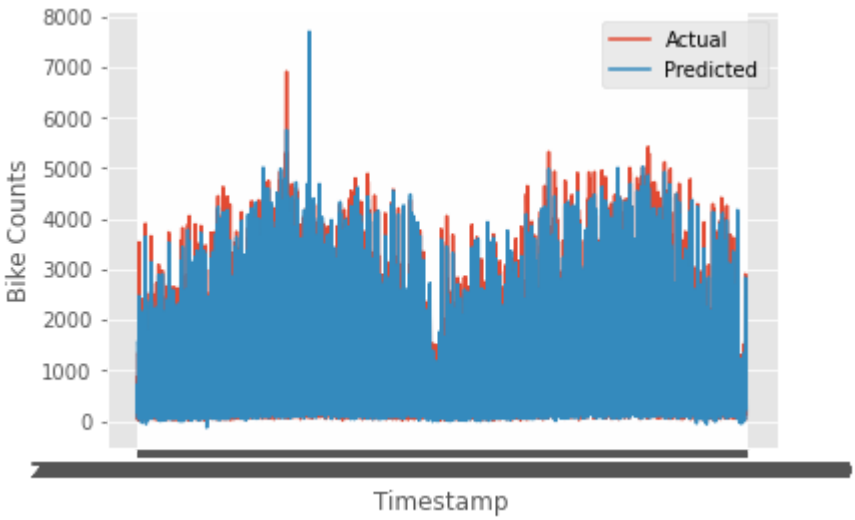

Figure 11: Scatterplot of actual and predicted bike counts

Some of the limitations of the model lie in the data processing and hyperparameter tuning that can be even more improved with even bigger dataset. Using a dataset for a duration exceeding two years can even improve the model. Further hyperparameter tuning, adding more features such as stations locations, or using a different algorithm like random forest model or other types of neural networks. Moreover, the performance of the model may have less prediction powers for other cities than London, where weather conditions have different effect and other weather features are needed to represent the weather such as thickness of snow in extreme cases where the threshold between acceptable weather conditions and extreme ones vary. In addition to that some other features that can be important for other locations might not be accounted for in this dataset and model, like holidays that differ between countries. Moreover the number of docks in station can also have an impact which influences the maximum number of bikes available in a station, and in paper [5], it was shown that the bike counts per hour is affected by the available bikes in stations in the previous hour, which is a good application of recurrent neural networks. Besides that, the testing data count is around 50% of the training data set, which suggests that bigger training data should be better for better accuracy, in addition to that more analysis to outliers of the test data can provide interesting insights about hidden features not accounted for in this dataset but that's outside of the scope of this project.

Conclusion

After interpreting the exploratory data analysis, a predictive model using regular neural networks through tensor flow package was created to predict range of bike counts in an hour. The features used in this model include: temperature, humidity, hour of the day, day of the month, month of the year, the year, weekend indicator, holiday indicator, weather code, season, & wind speed. In one of the models the final normalized root mean squared error of the model is 0.2058 and 0.25 for training data and validation data respectively, and in the best-fitting model the rmse for the training data was reduced to 0.1845. The model was later used to predict the bike counts in an hour for the test dataset. The mean of the predicted values is relatively similar to the mean of the training data with only 1.23% relative difference. Improvements to this model can be made by adding further features or performing additional hyperparameters tuning, however the predictive model obtained is satisfactory. This model can be used for scheduling redistribution shifts for bikes between stations in non-busy areas were demand is low and accuracy level is less critical given that the model predicts total number of bikes shared per hour in all London city, however for microscopic predictions, the model might not be as accurate given that demand per station might be modeled different and it can be influenced by different factors such the population distribution in the city and the land-use nearby stations.

References

1. **Research on the Forecast of Shared Bicycle Rental Demand Based on Spark Machine Learning Framework**
   Zilu Kang, Yuting Zuo, Zhibin Huang, Feng Zhou, Penghui Chen
   Institute of Electrical and Electronics Engineers (IEEE) (2017-10) https://doi.org/ghnjkh
   DOI: 10.1109/dcabes.2017.55

2. **Predicting public bicycle rental number using multi-source data**
   Fei Lin, Shihua Wang, Jian Jiang, Weidi Fan, Yong Sun
   Institute of Electrical and Electronics Engineers (IEEE) (2017-05) https://doi.org/ghnjkj
   DOI: 10.1109/ijcnn.2017.7966030

3. **Central Station Based Demand Prediction in a Bike Sharing System**
   Jianbin Huang, Xiangyu Wang, Heli Sun
   Institute of Electrical and Electronics Engineers (IEEE) (2019-06) https://doi.org/ghnjkk
   DOI: 10.1109/mdm.2019.00-38

4. **London bike sharing dataset** https://kaggle.com/hmavrodiev/london-bike-sharing-dataset

5. **Modeling bike counts in a bike-sharing system considering the effect of weather conditions**
   Huthaifa I. Ashqar, Mohammed Elhenawy, Hesham A. Rakha
   Case Studies on Transport Policy (2019-06) https://doi.org/ghnjkg
   DOI: 10.1016/j.cstp.2019.02.011

6. **A complete guide to the random forest algorithm**
   Built In
   https://builtin.com/data-science/random-forest-algorithm

7. **Machine Learning Crash Course**
   Google Developers
   https://developers.google.com/machine-learning/crash-course