CAPSTONE PROJECT

SEOUL BIKE SHARING DEMAND PREDICTION

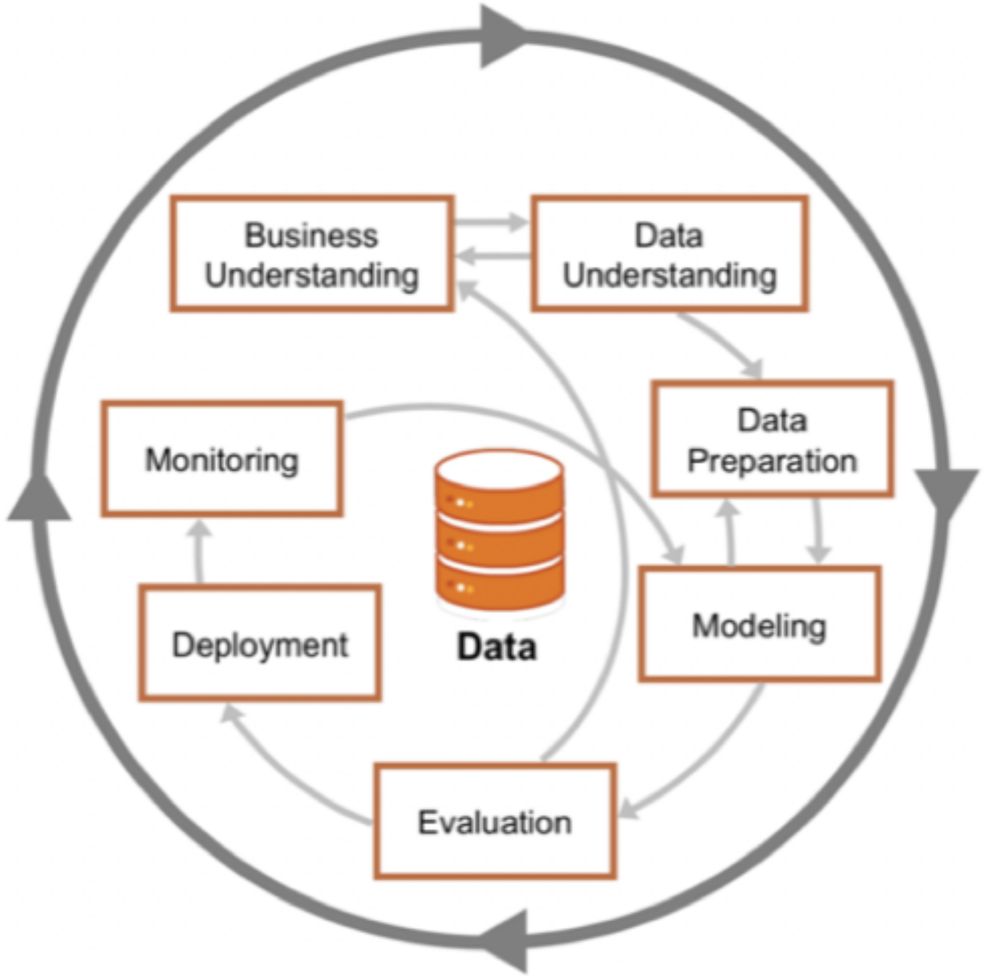
Business Problem:

Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required for the stable supply of rental bikes.

Data Science Process

We used CRISP-DM process for our project:

- Business Understanding
- Data Understanding
- Data Prepration
- Modelling the data
- Interpreting the results and conclusion



Business Understanding

Seoul Public Bike is an unmanned rental system that can be conveniently used anywhere, anytime by anyone. This bike sharing system allows people to borrow a bike from a “dock” which then can be returned to any other dock in the city.

The system was designed to resolve issues of traffic congestion, air pollution, and high oil prices in Seoul, and to build a healthier society while enhancing the quality of life for Seoul citizens. Hence it is important to predict the demand of the bikes to ensure these are readily available to the public.

The council is unable to forecast the demand of these bikes required in the city. There is a need for a model to predict the demand of these shared bikes. The business needs this model to build a strategy to meet the demand levels, or else if not met, people may not prefer to use this mode of transport. Our project aims at predicting the demand of the rental bikes for Seoul Council.

Data Understanding

- This dataset represents number of bikes rented from Dec'17- Nov'18 for each hour. (365 days)
- This contain 8760 rows and 14 columns.
- Three categorical features 'Seasons', 'Holiday', & 'Functioning Day'.
- One 'Date' variable
- Numerical type variables such as temperature, humidity, wind, visibility, dew point temp, solar radiation, rainfall, snowfall which tells us the environmental conditions for that particular hour of the day.

	Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0	Winter	No Holiday	Yes
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0	Winter	No Holiday	Yes

Feature Summary

- Date : Year-Month-Day
- Rented Bike Count - Count of bikes rented at each hour (Dependant variable)
- Hour - Hour of the day
- Temperature - Temperature in Celsius
- Humidity - %
- Wind Speed - m/s
- Visibility - 10m
- Dew point temperature -Celsius
- Solar radiation -MJ/m2
- Rainfall -mm
- Snowfall –cm
- Seasons -Winter, Spring, Summer, Autumn
- Holiday -Holiday/No Holiday
- Functional Day - NoFunc(Non Functional Hrs),Fun(Functional Hrs)

Data Preperation

Steps we have taken during this step:

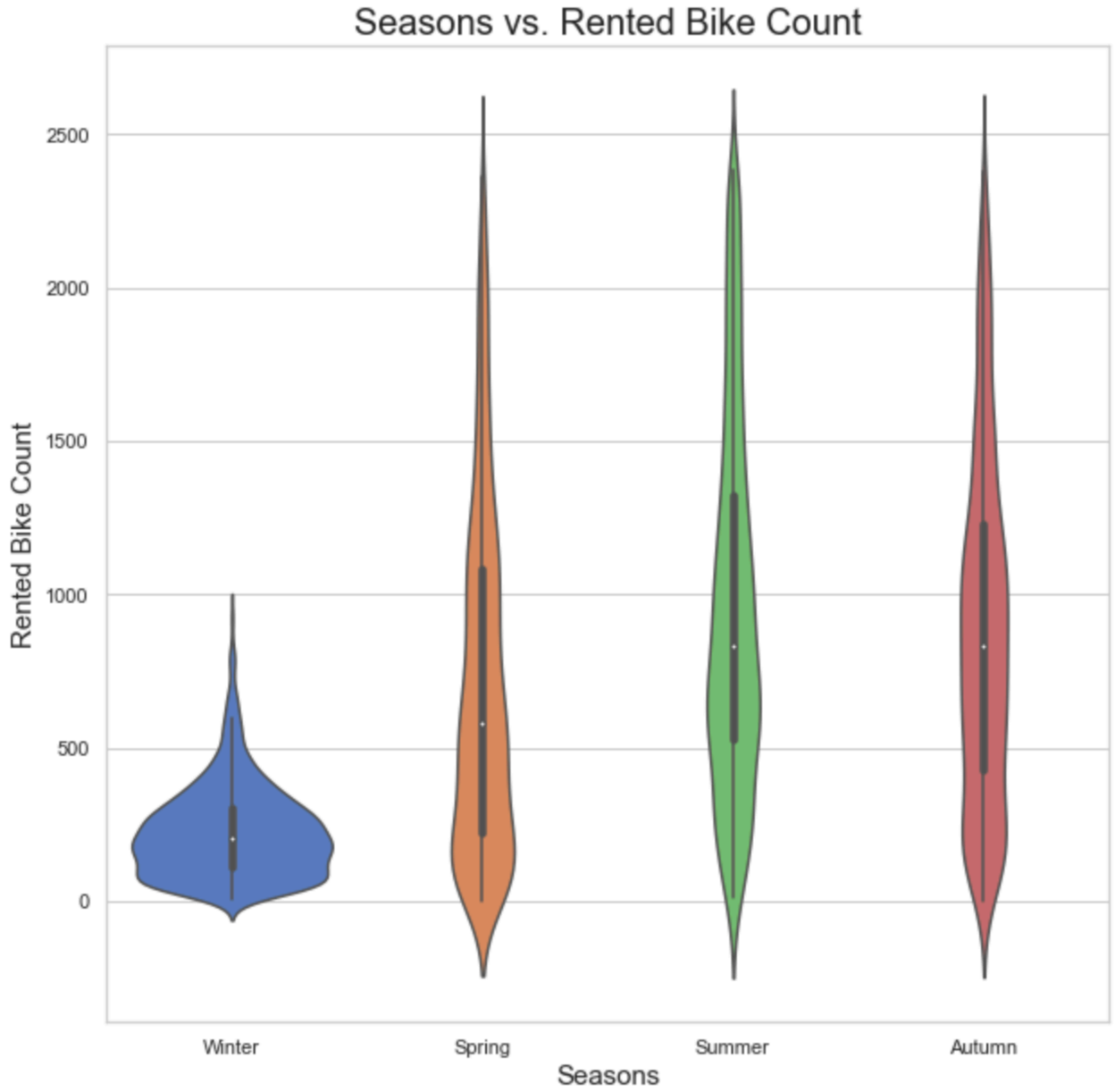
- Checked Missing values, Duplicate values or Null values in the data.
- Deleted unnecessary columns - Functioning Day as there are no bikes rented on the non- functioning day.
- Added new categorical features to explore the data further:
  - Weekend (Including Saturday and Sunday) - Yes/ No
  - Timeshift - based on the time intervals- it has Day, Evening and Night
- Dealing with the Outliers in Bike Count, Wind, Solar Radiation, Rain and Snow by capping with IQR limits
- Checking and removing Multicollinearity - we dropped Dew Temp column
- Converting categorical values using Label Encoder
- Normalised the dataset using Min Max Scaling

Relationship between Seasons and Rental Bike Count

We observed the relationship between the seasons and the rental bike count. As expected, the Winter season observed the lowest count while there were the highest numbers of bikes rented during the Summer season.

Whereas Autumn was the second highest busiest season.

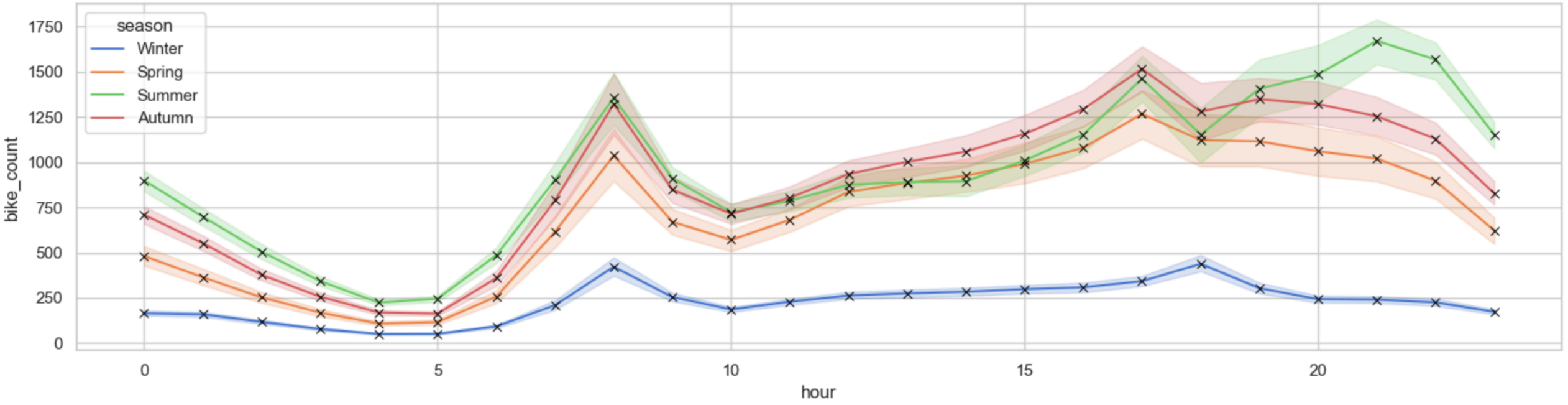
season	bike_count
Summer	2,098,149.0
Autumn	1,713,936.0
Spring	1,543,304.0
Winter	487,169.0



Seasons per hour Rental Bike Count chart

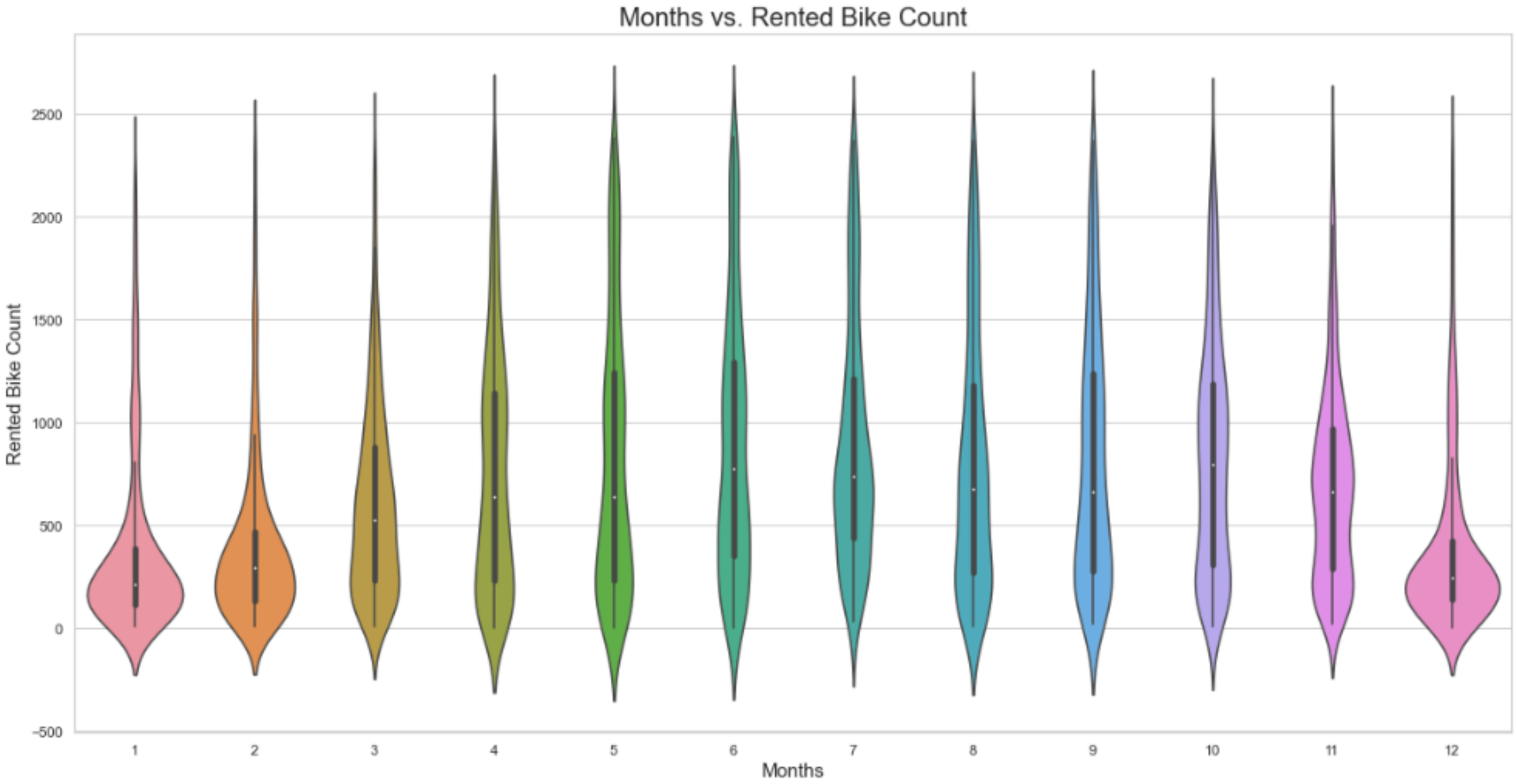
Further analysis between season and hourly bike count data shows there are particular hours of the day when the demand is high.

- We see a big spike from 7am-9am, this could be due to peak office time.
- And again from 3pm-8pm. This again could be due to a mix of reasons: peak working hours, family time, personal reasons.
- During Summers, we see there is more demand for the bikes late evening from 8pm-midnight.
- The demand in Winters is fairly low as compared to other seasons.
- The time from midnight to 6am is very slow with the renting of the bikes, so the council can plan any repairs or testing during these times.



Months vs Rental Bike Count chart

Summer : Jun- Aug	Winter : Dec-Feb
Autumn : Sep- Nov	Spring : Mar- May



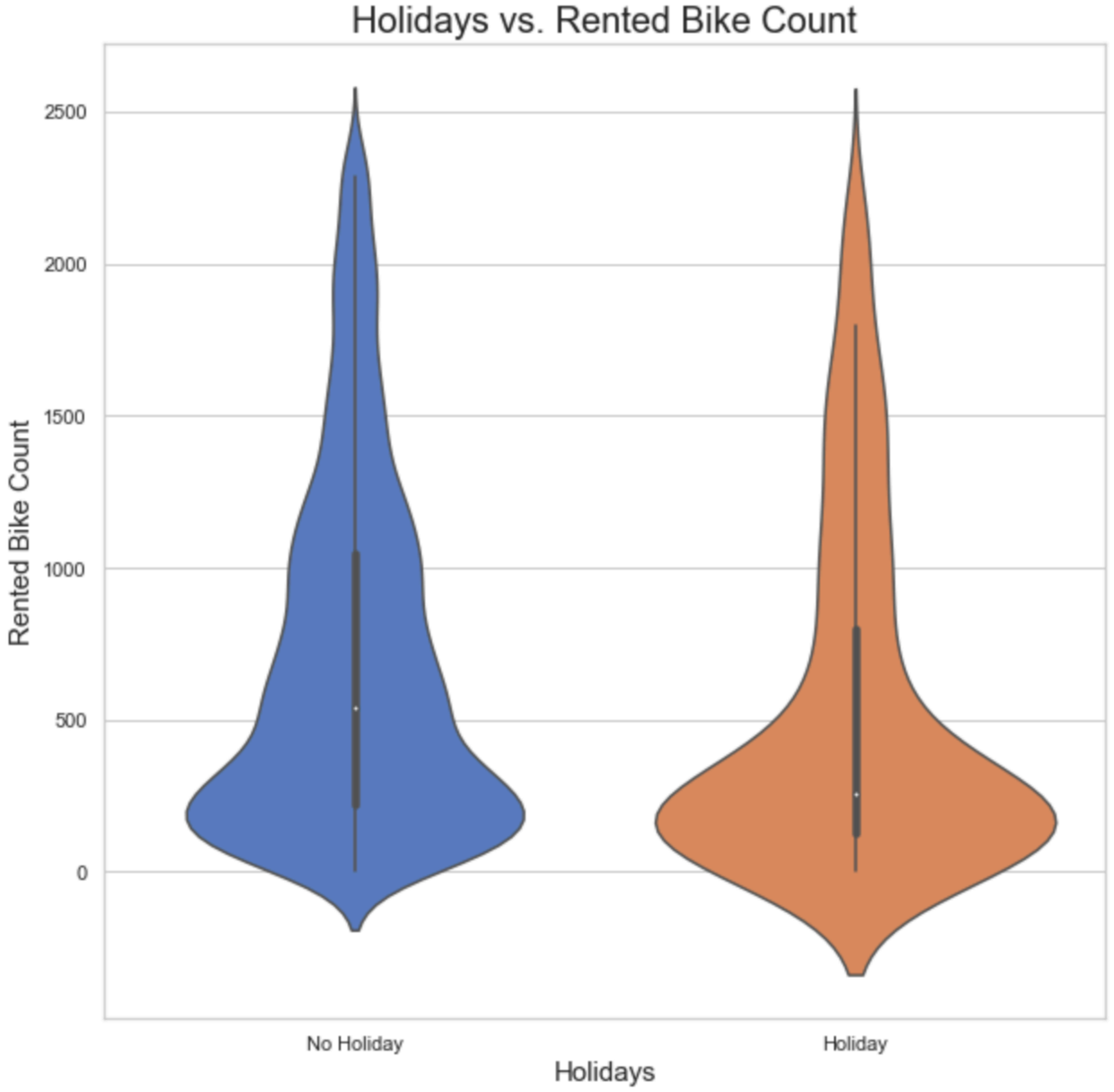
Relationship between Holidays and Rental Bike Count

It was interesting to see a low demand for the bikes on the days when it's a Holiday in Seoul. We see a drop in the overall count of the bikes rented out on the holidays as compared to non holidays.

The per hour average count of rented bikes on holiday days is 496 while on non holiday days is 675 which is 26% higher.

Total number of bikes rented:

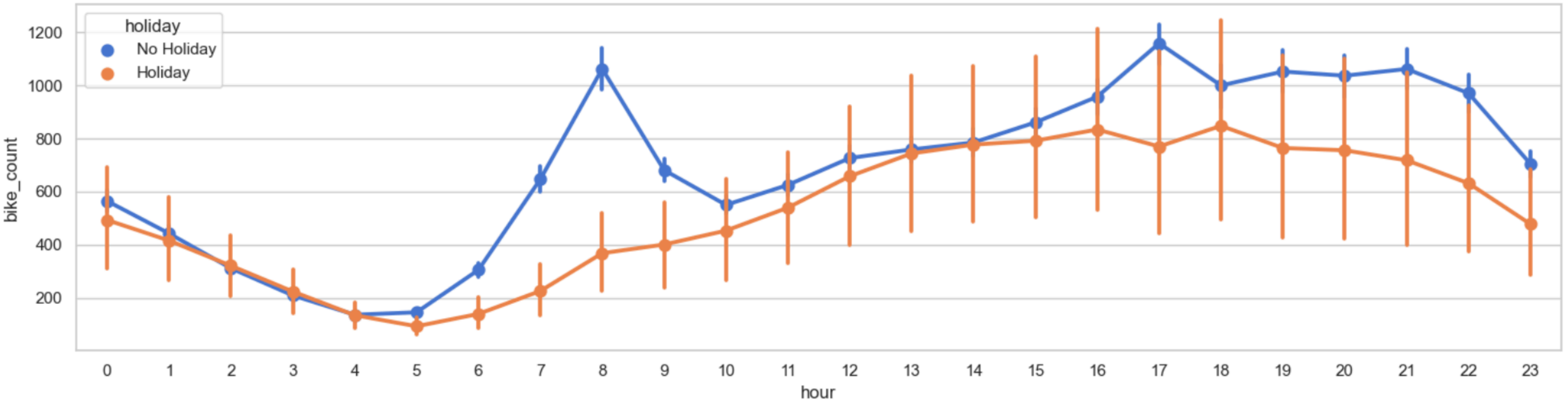
holiday	bike_count
No Holiday	5,628,521.0
Holiday	214,037.0



Holidays per hour Rental Bike Count chart

We see from the below chart

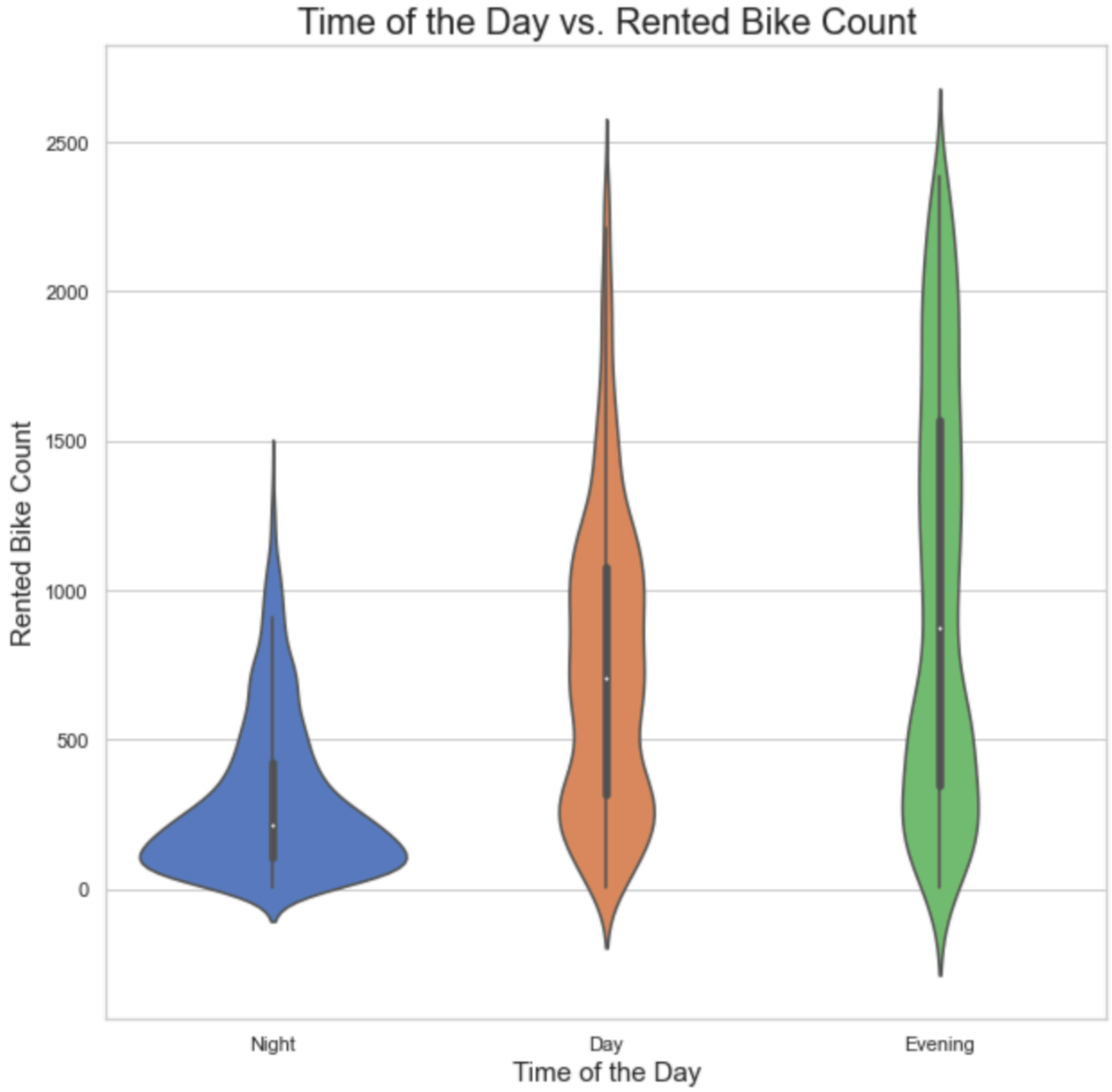
- During working days the demand is high between 7am -9am. While in the afternoon we see on an average more than 1000 bikes are rented out between 4pm - 10pm.
- On the holidays, even though the overall demand is fairly low, between 2pm-8pm the maximum bikes were rented out.



Relationship between Time of the day and Rental Bike Count

As expected, maximum bikes have been rented out during the day, followed by the evening. The night times are usually the slow times.

timeshift	bike_count
Day	2,670,448.0
Evening	2,430,487.0
Night	741,623.0



Modelling:
Linear Regression Model results

Parameters	Results
R^2	0.665
Adj R^2	0.664
Train MSE	0.019
Test MSE	0.018
CV Mean	0.020

OLS Regression Results

Dep. Variable:	bike_count	**R-squared:**	0.665
Model:	OLS	**Adj. R-squared:**	0.664
Method:	Least Squares	**F-statistic:**	1862.
Date:	Tue, 08 Nov 2022	**Prob (F-statistic):**	0.00
Time:	20:50:01	**Log-Likelihood:**	4801.6
No. Observations:	8465	**AIC:**	-9583.
Df Residuals:	8455	**BIC:**	-9513.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2837	0.008	33.821	0.000	0.267	0.300
hour	-0.2642	0.015	-18.117	0.000	-0.293	-0.236
temp	0.5868	0.009	67.350	0.000	0.570	0.604
humidity	-0.2298	0.010	-24.101	0.000	-0.249	-0.211
solar_rad	0.0208	0.007	2.904	0.004	0.007	0.035
rain	-0.4289	0.012	-34.608	0.000	-0.453	-0.405
season	-0.1319	0.004	-29.779	0.000	-0.141	-0.123
holiday	-0.0553	0.007	-7.919	0.000	-0.069	-0.042
timeshift	0.4112	0.011	36.051	0.000	0.389	0.434
month	0.0269	0.005	5.524	0.000	0.017	0.036

Omnibus:	79.236	**Durbin-Watson:**	0.823
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	129.830
Skew:	-0.020	**Prob(JB):**	6.42e-29
Kurtosis:	3.605	**Cond. No.**	20.4

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Conclusion

- The most important feature of the model is the temperature for our Linear model. The demand is high during the hotter days.
- Seasons and months have a strong impact on the demand for the bikes. Summer and Autumn months have higher demand. Winter is the month with the lowest demand for the bikes.
- Time of the day is also important factor to consider while planning. The demand is higher during certain times of the day during different seasons. 7am-9am and 3pm-8pm is the busiest time.
- Rainy days have a negative impact on rental bikes as we could have imagined. People dont prefer to use bikes when it’s pouring.
- Non holiday days are busier for rental bikes as compared to holidays.

Thank you!

Email: taniachopra01@gmail.com
GitHub: @TaniaChopra01
LinkedIn: linkedin.com/in/username/taniachopra01