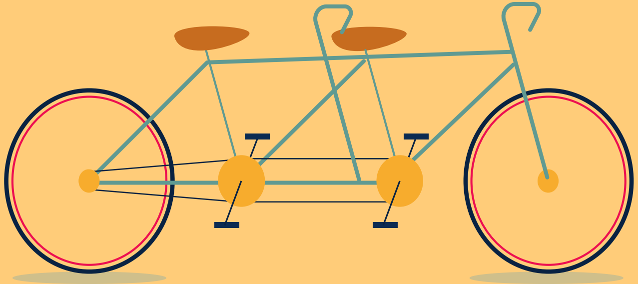
BIKE SHARING DEMAND - TIME SERIES ANALYSIS

TEAM-5



Our

Team Members

Team-5

Niranjan Reddy Masapeta

Purna Sai Mahesh Goud Palagani

Sujan Rao Chikkela

Utsav Savaliya

Abstract

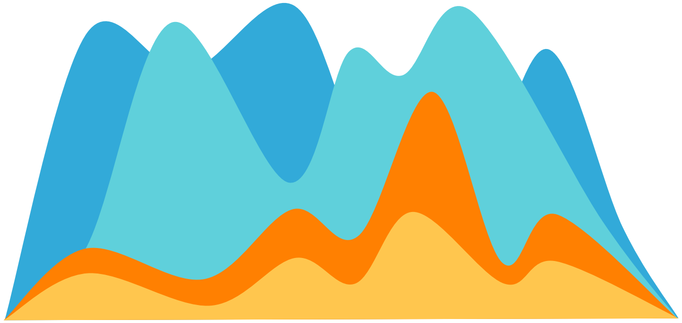
Time series analysis is one of the most popular data analysis methods from long time. In this project, we focus on bike sharing demand forecasting problem which is an application of time series modelling.

Introduction and Problem Statement

- Time series forecasting in statistics is a method where we predict variables that change over time.
- In our problem, we have a time-series data related to bike rental service where customers rent a bike for the day and we need to forecast the demand i.e. the number of customers who will rent the bike depending on the conditions around him.
- Our goal of the project is to forecast the number of customers expected to rent bikes given the previous data.
- Our goal includes applying different data analysis and visualisation techniques taught in this course to gain insights from the data.

Data Analysis

- Before jumping into the data modelling, we have to perform Exploratory Data Analysis (EDA) on the dataset to gain insights from the data.
- Data Analysis and Data Visualization are part of EDA methods.
- EDA is one of the most important steps in Machine Learning as the accuracy and efficiency of a Machine Learning model depends on the data the you feed to the model.
- EDA is also important because of the insights you gain from the data can be very helpful for the growth of certain businesses, in our case being Bike rental sharing business.

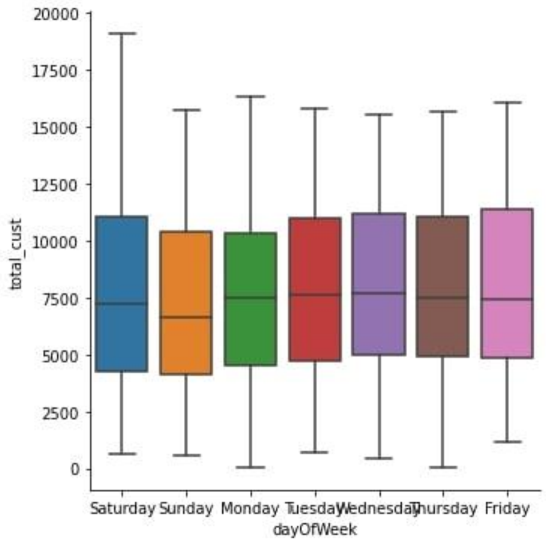


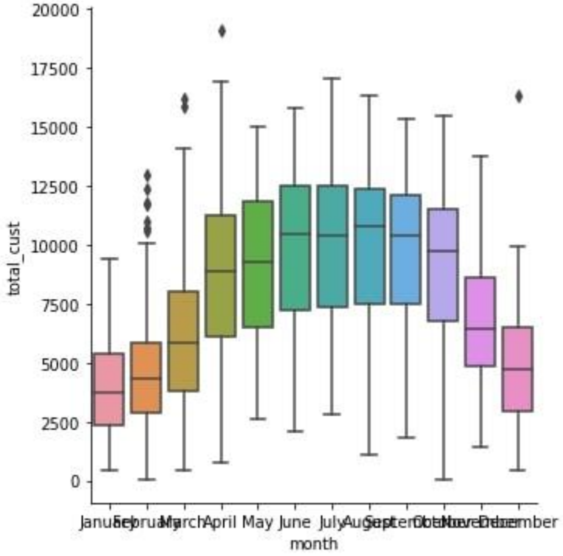
Dataset Overview

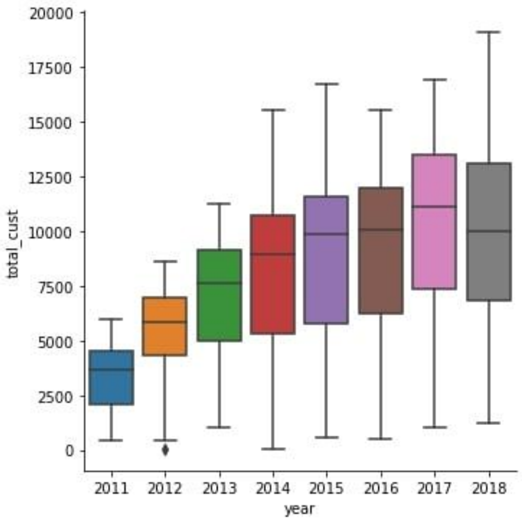
```
df_v2
```

	date	temp_avg	temp_min	temp_max	temp_observ	precip	wind	casual	registered	total_cust	holiday	rain	fog	ice	datetime	year
0	2011-01-01	5.20	-1.57	11.97	2.77	0.07	2.58	330.0	629.0	959.0	0.0	1	1	0	2011-01-01	2011
1	2011-01-02	7.34	0.88	13.81	7.33	1.04	3.92	130.0	651.0	781.0	0.0	1	1	0	2011-01-02	2011
2	2011-01-03	2.01	-3.44	7.46	-3.06	1.88	3.62	120.0	1181.0	1301.0	0.0	0	0	0	2011-01-03	2011
3	2011-01-04	-0.66	-5.96	4.64	-3.10	0.00	1.80	107.0	1429.0	1536.0	0.0	0	0	0	2011-01-04	2011
4	2011-01-05	0.91	-4.29	6.11	-1.77	0.00	2.95	82.0	1489.0	1571.0	0.0	0	0	0	2011-01-05	2011
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2917	2018-12-27	3.50	-3.59	9.12	-1.06	0.02	2.10	1150.0	4280.0	5430.0	0.0	0	1	0	2018-12-27	2018
2918	2018-12-28	8.23	0.61	11.21	8.09	16.84	2.00	166.0	1959.0	2125.0	0.0	0	1	0	2018-12-28	2018

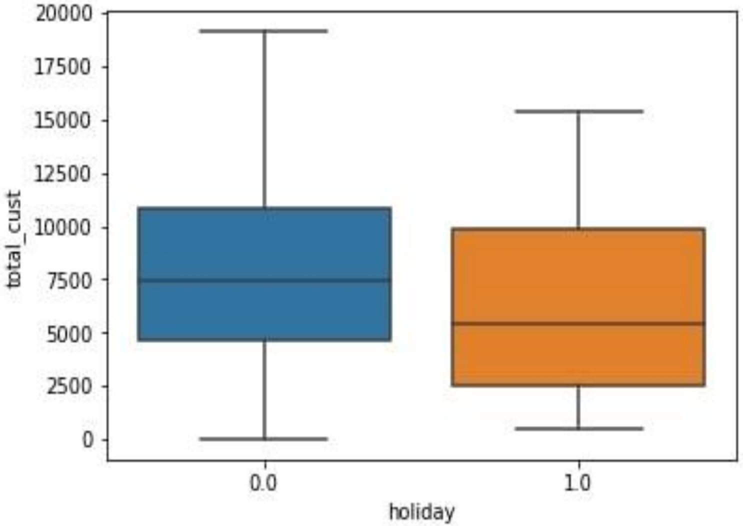
Data Analysis - Insights

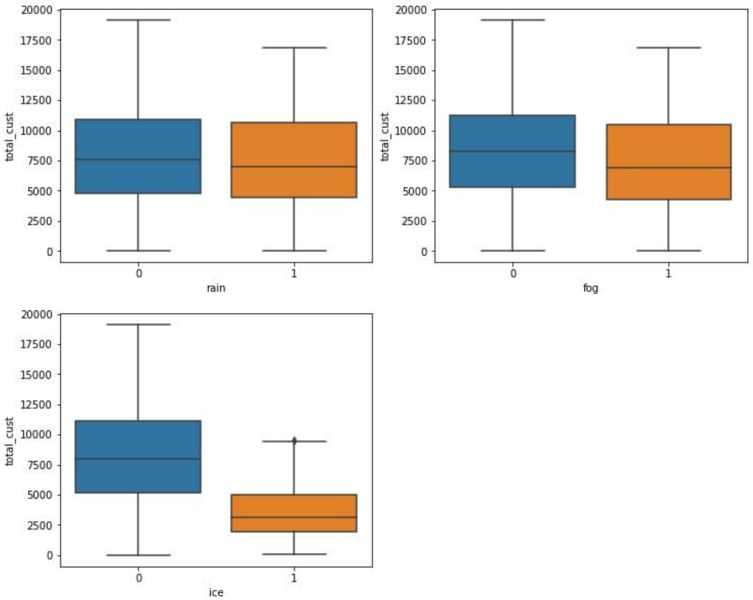




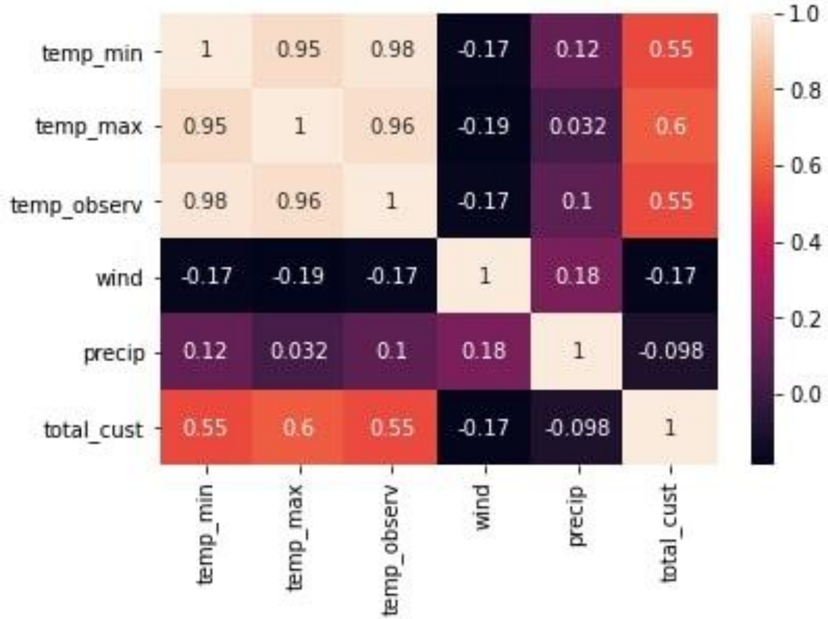


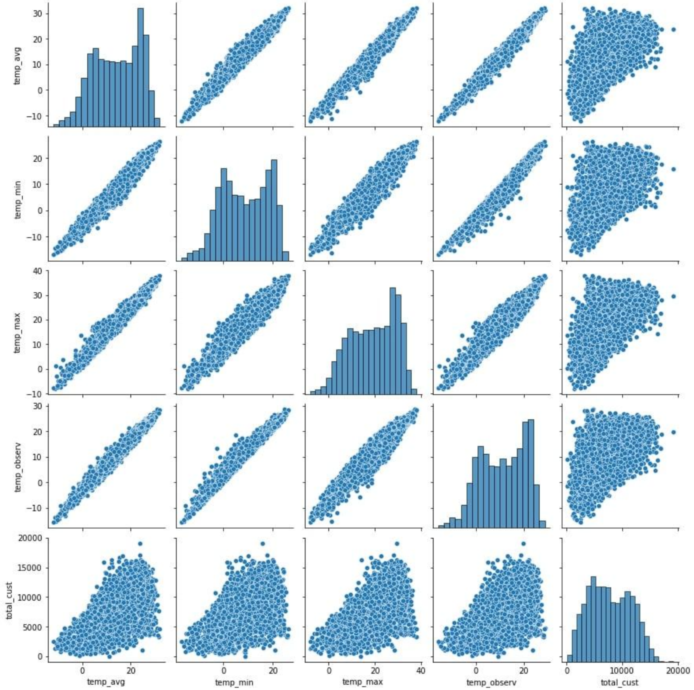
Data Analysis - Insights





Data Analysis - Insights

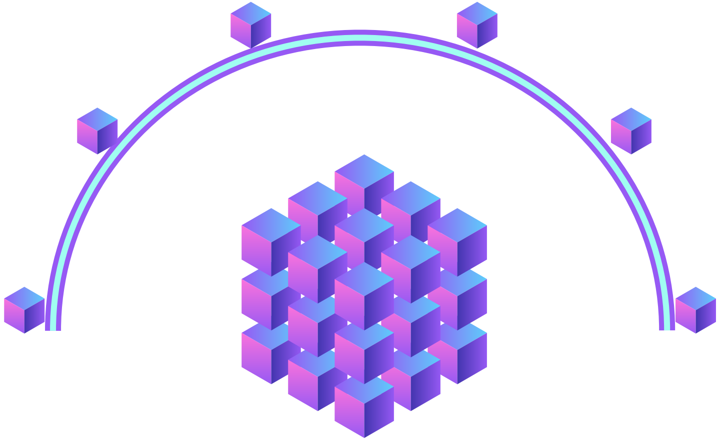




Data Modelling.....

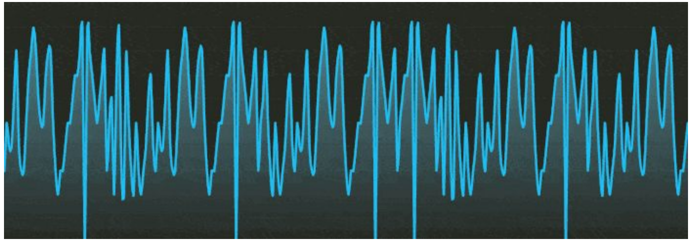
Before going to data modelling, our data is a time series data.

We need to check for stationarity



Time series analysis - Stationarity vs Non-stationarity

- A stationary time series is a series whose properties such as mean, variance and autocorrelation doesn’t change over time. If the properties change over time, then it is non-stationary series.
- We should ensure that the series is stationary in order to forecast correctly and if the series is non-stationary, we convert it to stationary.
- Stationarity is preferred because if the mean and variance increase over time, it is difficult to forecast as the series is changing over time.
- Augmented Dickey Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test are used to evaluate if a series is stationary or non-stationary.



Time series analysis - Stationarity vs Non-stationarity

- ADF test has hypothesis testing for difference stationarity whereas KPSS test looks for trend stationarity in the series.
- Non-stationary series is converted to stationary series using following methods,
  - Differencing: Removes series dependencies on time
  - Detrending: Removes trend effects from the dataset
  - Transformation: Converts into stationary using log transfer method
- After converting the dataset to stationary series, next step is data modelling.

Data Modeling and Results

ML Models	RMSE	COD	MAE
Ada Boost	0.37916392388199877	0.9999999897855025	0.06063569682151589
Gradient Boost	24.53625508101554	0.9999572260315428	18.720790303863236
XG Boost	11.893188350746476	0.999989950149935	9.00609441631289
Bagging Estimator	3400.463493643635	0.17845684502626363	2881.978636400845
Cat Boost	27.363339056767632	0.9999468012812855	21.775526527822716

THANK YOU!

Any Questions?