Project Proposal

**Project Title:**

1. **Bike sharing system**

* A bike sharing system is a service in which bikes are made available for shared use to individuals on a short term basis for price or free.
* Bike share systems allow people to borrow a bike from a "dock" which is usually computer-controlled wherein the user enters the payment information, and the system unlocks it. This bike can then be returned to another dock belonging to the same system.

1. **Problem statement and business use -case**

* A US bike-sharing provider BoomBikes has recently suffered considerable dips in their revenues due to the ongoing Corona pandemic. The company is finding it very difficult to sustain in the current market scenario.
* Company has decided to come up with a mindful business plan to be able to accelerate its revenue as soon as the ongoing lockdown comes to an end, and the economy restores to a healthy state.

1. **Data science work flow**

* We want to measure the counts of bikes shared based on temperature, windspeed, weekdays etc

Data provided us has factors like

1. instant: record index
2. - dteday : date
3. - season: season (1: springer, 2: summer, 3: fall, 4: winter)
4. - yr : year (0: 2011, 1:2012)
5. - mnth : month ( 1 to 12)
6. - hr : hour (0 to 23)
7. - holiday: weather day is holiday or not
8. - week day: day of the week
9. - working day: if day is neither weekend nor holiday is 1, otherwise is 0.
10. - weathersit :

- 1: Clear, Few clouds, Partly cloudy, Partly cloudy

-2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

- 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain +Scattered clouds

- 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

1. - temp : Normalized temperature in Celsius. The values are divided to 41 (max)
2. - atemp: Normalized feeling temperature in Celsius. The values are divided to 50 (max)
3. - hum: Normalized humidity. The values are divided to 100 (max)
4. - windspeed: Normalized wind speed. The values are divided to 67 (max)
5. - casual: count of casual users
6. - registered: count of registered users
7. - cnt: count of total rental bikes including both casual and registered

And we want to predict the count of bikes

* Data shared with us collected over a span of 2 years (2018 and 2019) across different seasons (springer, summer, fall, winter) and temperature is recorded.
* We want to build multiple linear regression model to predict demand of bike sharing over the seasons

We want to analyse different factors that would be important for prediction of bike demand in American market.

Also, how these factors impact bike sharing.

* First solution is to identify the factors that are significant in predicting the bike demand

Second solution want to deliver and build a model to predict demand based on these factors.

1. **EDA questions**

* How bike sharing demand affects on different seasons
* Is bike share have dependency on week days and weekends
* Is bike sharing affects on temperature?
* Is bike share depends on holidays?
* How bike share affects on windspeed?

1. **Data Analytics**

* **Descriptive** analysis

Bike count has effect on registered users and temperature

* **Diagnostic** analysis

Registered users may find bike sharing facility as the means of transport to their office.

When temperature is too low,ie.winter season people may find difficulty in travelling by bike.

* **Predictive** analysis

Increasing registered users, can increase bike sharing demand.

* **Prescriptive** analysis

Registered users can give some offers.

Try to sanitise vehicles after every use.

1. **How our problem can be solved?**

* Increasing count of registered users by give some offers during subscription
* Try to sanitise bike after every use
* Increase the count of pick -up-point, in every junction(Near bank, IT-hub etc)
* Increase the count of bike in morning & evening (office-time)

1. **Who will be impacted?**

Customers will be highly impacted by these system, If it function in proper way. So that revenue will increase.`

1. **Data collection**

Dataset-is in the form of excel sheet

Dataset in the form of excel file.

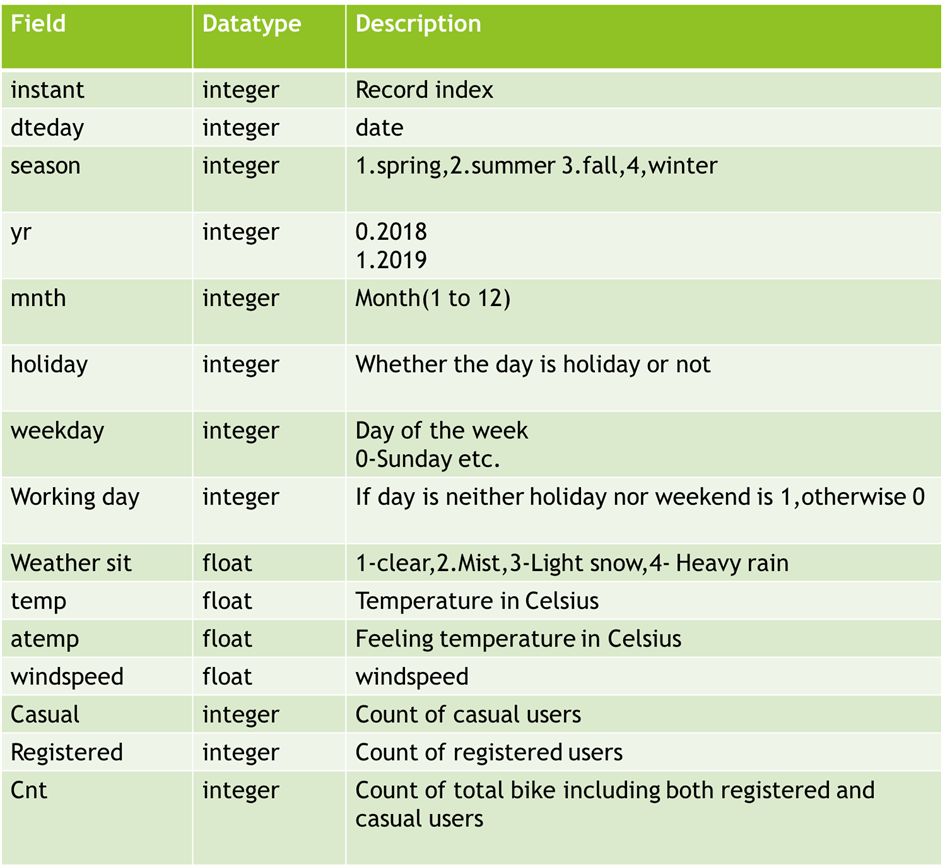
Data set consists of data of span of 2 years(2018 & 2019)

Dataset has 730 rows × 16 columns

Dataset file is in the form of.csv

It has size of 56.1Kb

Datatypes



**9. Data Pre-processing ,preparation and feature engineering**

* Identifying data quality-
* Check for data inconsistency, all datatypes are correct and no encoding issues.
* Handling missing values
* Treating categorical values
* Treating outliers
* Do scaling
* Feature engineering

**10. Machine Learning and Model Selection**

Multiple linear regressioyjmzn