**Contribution**

Environment has been a main topic of discussion in summits and conference across the world. The recent World Environment Summit in Paris has emphasized a lot on the saving the environment. With the rapidly growing economy and increase in the spending power of the people there are a number of vehicles on the road. This has increased the air pollution exponentially. The increase in air pollution has decreased the content of the ozone layer in the environment. The ozone layer is one of the important component of the atmosphere. The ozone layer surrounds the atmosphere which protect the earth from the ultraviolent rays which is emitted by the sun. The ultraviolent rays emitted from the sun will increase the heat in the earth atmosphere which will result in the phenomenon called Global Warming. In the recent years it is noted the global warming has been damaging the environment drastically. The global warming has damaged a number of ice glaciers, increase in the level in the ocean which has given way to a number of natural calamities. Hence immediate attention needs to be given to the ozone content in the atmosphere. The world today is also building a smart cities to ensure proper monitoring of the city’s assets. The city’s assets includes the transport facilities, pollution control, parking arrangements, traffic updates, entertainment planning etc. Smart cities are a good idea in the give scenarios which will help us understand the areas which require our attention. The smart city helps the city administrators to solve the issues of the citizens in an effective and transparent way. To understand the various aspects of a smart city we study the various department data of Aarhus a city in Denmark which is termed as a smart city. We picked the parking data and the pollution data to understand the impact of a proper planning of parking infrastructure and the information about the pollution index in the world.

It was observed during our research it was observed that sixty five percent of the air pollution is caused by the search of parking by the vehicles on the streets. It means that the vehicle spends sixty percent of its time in search of parking. Hence we have analyzed the parking information along with the pollution index to understand impacts of the effective parking on the ozone layer. As the vehicle count increases on the road there carbon monoxide is emitted from the vehicles effect the environment. The carbon monoxide when induced into the environment will reduce the ozone content in the atmosphere. This will deplete the ozone content in the atmosphere which will in turn affect the environment adversely. This depleted ozone layer in will increase the heat radiation on the surface of earth. This increased heat radiations will cause a lot lead to a lot of natural calamities .Hence we try to study the impact of the parking information and its effective planning using the parking information. Later we used to the concept of time series to predict the effects of the vehicle counts and the effective parking methods. Our contribution will help the pollution control department, transport and the traffic department to ensure that they manage the congestion in a junction and will also allow the pollution control departments will be able to make better arrangements to control the pollution.

**State of-the-Art**

There has been a lot of research and work in the area of parking analytics today. In today’s scenario where the population increase across the world and booming economy is making a lot of people migrate from the rural environment to the urban set up. This migration asks the city environment to ensure that we have a proper planning for the city’s infrastructure specially parking. There has been a good amount of work will done to help build an efficient parking managing systems. The research work done so far can be classified into these five categories:

* Blind Search Parking Method
* Parking Information Systems(PIS)
* Reservation Based Parking Information System
* Buffer Parking Information Systems(BPIS)
* Smart Reservation Parking Systems.

Now we shall discuss each of these categories in detail and also identify the strength and weakness of each of these categories.

**Blind Search Parking Method**: In this type of parking method the driver drives around the parking slots to find an empty parking slot. This is one of the naïve methods employed in many cities across the world. This will not reduce the air pollutions in the cities. This will in turn contribute the increase in traffic congestion and the pollution. The traffic congestion will lead to many complications in planning the city’s infrastructure. Hence the researchers worked on creating a better parking information systems which will help the drivers identify the parking spots. This is very general parking methods which are useful in the area which are smaller. This will contribute the number to a number of problems in the environment.

**Parking Information Systems:** A survey done in the city of Aarhus for cycle parking where a chip was inserted in the bicycles to identify and calculate the amount of time it takes to find a parking spot. This chip is connected to the phone. Which gives to the biker the number of available parking spots in the area. This was improved to the motor vehicles. The information screen set up in various places in the cities. Will give the driver a total number of parking spots available. The information is helpful to the driver to decide where to park the vehicle. This will lead to the same blind search parking method where the driver search for the empty parking spot. There are two main drawbacks in this method. They are 1.There is no information about the filled parking spot which will reduce amount of time the driver spends in looking for an empty parking spot. And also this will still have the same impact on the atmosphere. The search for parking will still increase the pollution in the air.

**Reservation Based Parking System:** The next level of parking systems included in devising a database which will have the information of all the parking spots in the city. These systems communicate with the citizens through a smart phone application. This application gives the information about the number of parking slots available in the area where the driver is looking for. The application will later allow you the driver to book the parking slot for the driver. This was a major improvement in the parking information system. But there are a number of drawbacks in this system. The system does not have an upper limit for reservation on each parking. The system is not intelligent enough to understand the number of reservations it can take. Secondly once the parking spot is reserved the other drivers don’t have access to that parking slot. The remaining drivers will retort by to the blind search parking method to search of an empty parking spot.

**Buffer Reservation Based Parking System:** To overcome the drawbacks of the previous systems. The reservation based parking systems architects came up with the concept of Buffer Reservation Based Parking System. The system will act in a similar way in as the normal Reservation Parking System works. The only difference will be the system will not disclose decided number of parking slots until the necessity for it arrives. These undisclosed parking slots are called buffer slots will be release the need for it arises. The number of undisclosed parking slots are called the buffer parking slots. The main advantage of this system is that we can understand parking behavior patterns in the area and also will be helpful in identifying the traffic congestions in the area. The drawbacks of these systems are 1. We will not know how long it will take the driver to get the empty parking slot. We will not the amount of time we will need to wait to release the buffer parking slots we have kept. 2. Once the parking slot is reserved the slot is not available for the other drivers. This will again go back to our blind search method which will again impact the environment.

**Smart Reservation Parking Systems**: In the Smart Reservation Parking Systems we have the similar to the Reserved Parking System. In this scenario we have the data bases connected to the smart phone application which will give the driver the number of parking slots available to the driver. Once the parking slot is reserved by the driver a quick response code is sent to the driver’s mobile using the short message services. Once the driver confirms the reservation the driver has to enter the quick response code. Once the code is entered and confirmed the parking it will be reserved for a time period of forty minutes. During this period of forty minutes this parking is reserved for the driver only. After the forty minutes are expired the system will release the slot for other drivers. This system solves a lot of problems which the other projects have not attempted to resolve.

All the above parking systems have contributed in developing an efficient parking systems. These systems in collaboration with big data has given the city administration a heads up on the various parking behavior patterns. This also helped the transport administration to regularized the parking tickets and manage the traffic congestion by adding additional workforce to ensure that there is a proper flow of traffic. These systems can become the trendsetters we can use the data to build better infrastructure.

**Data:**

The data we used for our analysis are 1. Parking and 2. Pollution. The parking data consists of the following data fields. The data is for the period of 12 months from May 2014 to October 2015 .They are

* Vehicle count
* Update Time
* Total Space
* Garage Code

**Vehicle count**: This gives us the information about the number of vehicles which can be parked in the garage.

**Update Time:** This gives the time the vehicle enters the given parking slot.

**Total Space:** The total spaces available in the parking slots.

**Garage Code:** This gives the unique garage code which will help us identify the garage where there cars are parked.

The second data set used was the pollution data. The data is a collection of pollution measurements designed to complement the vehicle traffic dataset above. For the pollution mockup stream we decided to simulate one sensor for each of the traffic sensor at the exact location of this traffic sensor. We have provided values for carbon monoxide, nitrogen dioxide, sulphur dioxide, particulate matter and ozone index levels according to the Air Pollution Index.

The data consists of the following columns

1. Ozone
2. Particulate matter
3. Carbon monoxide
4. Sulphur dioxide
5. Nitrogen dioxide
6. longitude
7. latitude
8. timestamp

The reason why the above the quantities impact the atmosphere the most. The stream generation works as follows: each sensor measurement (e.g. carbon dioxide) is initially assigned a value between 25 and 100. Every 5 minutes, the values will be updated as follows:

* if the value was below 20 before, it will now be the last value + random integer between 1 and 10.
* if the value was higher than 210, it will now be the last value - random integer between 1 and 10
* else the value will be last value + a random integer between -5 and 5

This way the measurements won't erratically jump between low and high values and represent a more realistic stream but still won't go out of bounds (unrealistically low or high values).

**Ozone**: Ozone is the one of the important part of the atmosphere. The ozone layer around the earth protects it from the ultraviolet rays produced. This has a cooling effect on the earth atmosphere which will help preserve the nature.

**Carbon Monoxide**: The carbon monoxide is the components which are emitted by the vehicles impacts the atmosphere by polluting it. This component will be affecting the ozone layers in the atmosphere. Hence this is the important parameter we need to emphasize.

**Sulfur Dioxide:** This is the chemical component which is toxic in nature which gives a pungent smell. This will be released by the industry waste and in a few places from the volcanic activity. This is a major air pollutant which impacts the health of living beings. This will lead other devastating natural calamity such as acid rain etc.

**Particulate Matter:** This is also known as particle pollution is made up of tiny solid particle and liquid droplets which are found in the air. This can be made of acids organic chemicals, metals and soil or dust particles. These might be manmade or natural**.** This will create many natural disasters.

**Nitrogen Dioxide:** Nitrogen oxide is a chemical compound which is an intermediate in the industrial synthesis of nitric acid, millions of tons of which are produced by the industries each year. This is a reddish brown gas which has a biting odor and is a prominent air pollutant. This gas is given by the industry which can pollute the environment. This has a number of effects on the health which will have created the number of ill health to most of the population around the area where the industry are.

**Longitude:** The longitude is the geographic coordinate that gives east and west position point on the earth’s surface. This along with the latitudes will give the exact position of the place we intend to go.

**Latitude:** Thelatitudeis the geographic coordinate gives the North and South position point on the earth’s surface. This helps us to get the accurate position where we want to go.

**Method**

The Linear **Regression** is algorithm we are using for the studying the impact of the ozone layers because of the delay in parking. Linear regression is a supervised machine learning algorithm. This algorithm has been a base for many more complex algorithms in machine learning. Linear Regression is an approach for modelling to establish a relationship between a scalar dependent variable and one or more explanatory variable. The dependent variable is usually represented as by Y and the explanatory variable are usually denoted by X. if the regression has only one explanatory variable then it is called Simple Linear regression. If the regression model has many explanatory variables then it is called Multiple Linear regression. In a linear regression explains the relationships between linear predictor functions where the unknown estimated parameters. Using final output of the prediction is always a straight line. Linear regression has been useful in many practical uses. Most applications fall into one of two broad categories:

* If the goal is prediction, forecast, or error reduction then the linear regression can be used to fit a predictive model on an observed data set on the dependent variable to the explanatory values. After developing a fit model more value can be given to ensure that the model works fine.
* The linear regression model can also be used to quantify the relationship between the dependent variable and the explanatory variable.

The linear regression model are often fitted using the concept of least square approach. They can also be fitted by minimizing the lack of fit.

There are a number of assumptions made to build a proper regression model. The assumptions are

* **Weak homogeneity:** This means that predictor variables or the explanatory variable can be treated as the fixed values rather than just random variables. This means that if we assume that the model is error free although this is not realistic we can make a conclusion we can predict the final output of the linear regression model free of errors.
* **Constant Variance :** This means that different response variable have the same variance in their same variance of errors regardless of the value of the predictor variables
* **Independence of errors:**  This assumption indicates that the errors of the response variables are not correlated with each other. This is a much stronger assumptions as any kind of correlation between the errors will make the prediction of the explanatory variable a bit difficult to build the final model.
* **Lack of Multicollinearity:** The lack of Multicollinearity in the explanatory variables will be make us assess the dependent variables will be in a way which we can ensure we get realistic results for an analysis.
* **Linearity:** This means that the mean value ofthe response variablein the linear combination of the parameter and the predicator variables. This is much less restrictive than the other assumptions.

Now the concept of the linear regression works on the estimation method to ensure that we have make predictions with maximum accuracy. A large number of procedure in machine learning the have been developed for parameter estimation inference of the results. The examples in this categories are the classification techniques and allocation methods. The estimation methods are of two types they are 1. Ordinary least square methods and 2. Generalized least square methods. As the accuracy of the linear regression model depends a lot on the least square parameter. The least square parameter has to be minimized to get the accurate model to forecast the dependent variable.

Now let us look at the estimation methods of the linear regression models in detail.

* **Ordinary Least Squares**: This is the simplest and the common estimator models to be used in the prediction of the dependable parameters in the linear regression model. This is very easy to compute and the easy to analyze. As the only action which we need to do here is to ensure that we minimize the least square errors to get accurate model for linear regression model.
* **Generalized least square**: This is an extension ordinary least square methods that allows efficient estimation of the beta values are different or are correlated in anyways. As long as the estimation values are not correlated the error terms will also be independent of each other.

The Linear regression model have found its applications in various fields such as biological, behavioral and social sciences to describe possible relationships between variables.

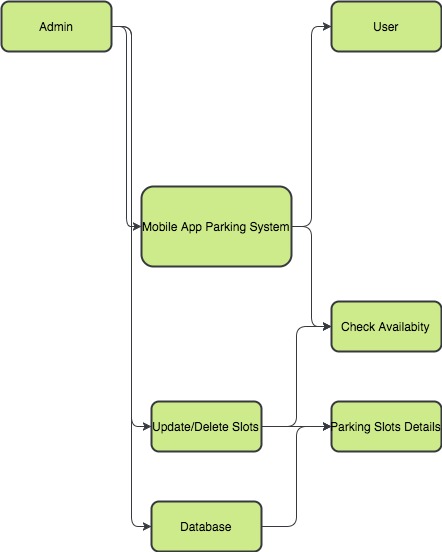
**Trend Lines**: A trend line represents the trends of the long term movements of the dependent variable after the explanatory have been accounted for. This gives us the exact predications of the dependent variables.

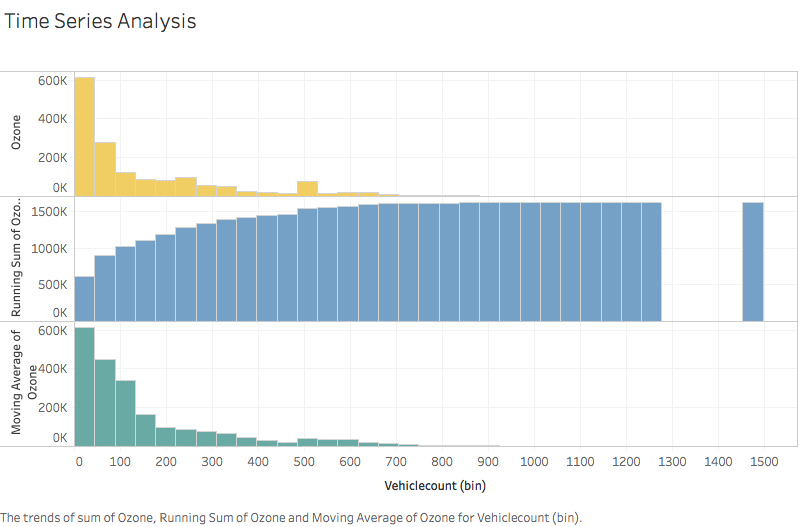
**Epidemiology:**  The early evidence which are related to tobacco smoking to mortality and morbidity came from the observational studies employing regression analysis.

**Finance**: The capital and asset price model uses linear regression as well as the concept of beta for analyzing and quantifying the risk of the model.

**Economics**: This is the one arena where the linear regression model is used as the predominant empirical tool in economics.

**Environmental** **Science**: Linear regression finds the applications in a wide range of environmental science applications. This was useful in the analysis of fish and benthic survey to measure the effects of pulp mill or metal mine effluent on the aquatic system.

Post applying the linear regression analysis on the parameters we use the time series analysis to build predict the effects of the delayed or inefficient parking systems on the environmental measures.



**Time Series Analysis**

A time series is a sequence taken at a successive equally spaced points in a time period. The time series analysis compares methods for analyzing times series data in order to extract the meaningful statistics and other characteristics of the data. Time series forecasting is used to predict future values based on the historical data we have. While the regression analysis is often employed in such a way that the current values are independent of the time series affect. A time series is a series of data particular data points indexed in time order. A time series is a sequence taken at successive equally spaced points in time. The fields which has motivated the time series analysis are quantitative financial analysis, communication engineering for signal detection and estimation. It further found its use in pattern recognition, clustering, machine learning, anomaly detection and forecasting techniques. The other advantage of the time series plot is that it can be used for panel data. The panel data is a general class of multidimensional data where we can implement the time series data set in a one dimensional data panels. A data set can exhibits both panel data and time series plot. Further when the information is transferred across time often we get a good time series plot. One of the classic example of the statistical Markov model is which the system model is assumed to be unobserved stats and yet we can make a process a time series plot to get the information.

The major goals of the time series plot are as follows:

* Identify patterns in correlated data that is trends and seasonal variations.
* Understanding and modeling the data
* Prediction of short term trend analysis
* We can also perform intervention analysis which says how a single event can change the time series.
* And also how a deviation from the point can indicate a problem in the analysis or the forecast.

Understanding the mechanisms of a time series will allows us to build a mathematical model that can explain way the prediction and monitoring can be done. The assumptions we make to build a time series model is that it follows a systematic pattern. The most common patterns are generally trends and seasonality. Trends time series is either linear or quadratic in nature. To find the trends average or regression analysis is often used. Seasonality is a trend that repeats itself systematically over time. The second assumptions is that the data exhibits a random process so it is sometimes hard to understand the systematic patterns within the data. Observations made over time will be either discrete or continuous. One of the most important property of the time series is that the observations are dependent on the previous observation. This phenomenon is called auto correlation. Hence this will lead to the challenging process of identifying the auto correlation.

Now there two main approaches for analyzing the time series. They are 1. The time domain and 2. Frequency domain. The analysis in the time domain is done is often used for stochastic observations. The frequency analysis are used periodic and cyclical observations. As already stated above that the time series data is correlated which means measurement are related to the previous observation and one change in the parameter might affect the entire analysis.

In our method we have the inculcated the previously implemented systems, we have the added the features of identifying the areas where the pollution is the highest and help them the pollution control department get a better equipment to ensure we have proper arrangements are made for controlling the pollution. Then we will be able to build a better environment. We latter then add the concept of cross validation. The cross validation will give us how much the model is accurate and how well we can predict the impact of the parking analytics will influence the ozone layer. The cross validation of a model is better than the residuals. In this method the data is divided into training data set and test data set. The training data set is where we perform and test various models which we intend to run. Once we run the required machine learning algorithms and run the model for accuracy to get the perfect fit. The details of each test is accurately studied and we pick the best model with better accurately. This will give us a varied option to apply a better model to implement with the data given. After testing the various models we identify the best model to the test model. We run the model on the test data and get the information and the final analysis we need. This will help us understand the data better and give us a better change of predictions.