1. **Abstract**

Increase in the economic behavior and upgrade of living standard the ratio of the people across the world there is an increase in the motor vehicles on the streets of all the major cities across the world. With the increase in the number of vehicles on the roads. There is an increase in the air pollution which is leading to many climate changes across the world. One of the main the reasons for the increase in the air pollution is the search for parking which increases the pollution in the atmosphere. In this paper we have tried to extend the Smart Parking System Based on Reservation to use to monitor the pollution in the city. Which will help us build a smart parking and smart environment. The analysis results show us the area where the pollution high which can help the administration build proper parking infrastructure.

**2. Work by Competitors**

There has been a number of the research work which has been done by the number of data scientists who have worked on parking analytics and various parking methods. We look at the various parking management systems and existing parking methods.

**2.1 State of Art Parking Management**

We discussed around the 1. Blind search parking 2.Parking Information System and 3. Buffered Parking Information system. We discuss this each of these methods in detail below.

1. **Blind Search Method**: Blind search method is a simple strategy applied for the users when there is no parking information. In this case the driver keeps driving till they find the parking within the certain distance to their destination. The drivers will not stop searching until they find a place. If the driver doesn’t find a space he will extend the search and continuously look for the parking near the place where he wants to go. This process has many drawbacks. This will can create a lot of traffic congestion in any location during busy hours.
2. **Parking Information System:** This is the mechanism is commonly adopted by the current state of the smart parking design. This system will publish the number of vacant spaces in the parking centers nearby. The driver will be able to decide their desired parking destination where the parking lot has available space accordingly to their need and the area where they are comfortable. However if the number of vacant spaces in a parking lot is very limited in busy hours. This will lead to a phenomenon called multiple car chasing single space. This will cause severe traffic congestion. As per the census it is stated that seventy percent of the traffic on the streets is because of the search for parking. This will increase the carbon monoxide in the atmosphere and increase the pollution. The increase in carbon monoxide will lead to the decrease in the ozone content in the atmosphere which will lead to the phenomenon called global warming. This has been common problem around the world. This will lead to the same problem as the blind search that the drivers will be driving around till they find a parking spots. This system too has many drawbacks. For example fake parking requests, driver identification and also lack of planning during busy hours will lead to the same problems as mentioned previously.
3. **Buffer Parking Information System:** The Buffer Parking Information Systems was an advanced version of the parking information system. The functioning of this is similar to the parking information system. The main difference between the two systems is that the system will ensure that system doesn’t release all the vacant parking spaces during the busy hours. These spots are called the buffer spots. The system will release the only a limited amount the vacant spaces during busy hours. Once all the available spots are filled the system will release the remaining the vacant spots for the drivers for parking. This system ensured to a certain extent that the lack of planning during busy hours is solved to a certain extent. But the fake parking request still exists and further the how long will the parking slot be available for the driver is not addressed.
4. **Smart Parking System based on Reservation:** This is the system we have the used to improve and add an additional leg to understand the pollution content in the atmosphere and then tally the same with the location and help the administration build a better infrastructure and measures to contain the pollution. The system work similar to the Parking Information System. The Smart Parking Information system has a database which has the connection to the phone through an application. The driver driving the area can log in to the application through its smart phone. The application will give the number of vacant spots near the areas where the driver is looking the vacant spots. Once the driver decides where he wants to park his vehicle he will be able to request for the spot using the application on his phone for his reservation of that parking spot. Once the reservation is confirmed the system will send a quick response code or QR code to the driver using the Short Message Services (SMS) when entered will reserve the parking spot for the driver. Once the reservation is confirmed the driver will get a message with the spot id on his phone which will help him park his/her vehicle. This system was an advanced version of the parking information system and the Buffer Based Parking Information System. The advantage of this system this will considerably reduce the fake parking request, driver identification and delay in reaching the parking spot. The reservation of the parking slot is valid for thirty minutes from the time the slot is reserved, post which the slot will be released for the next driver to use it.

**2.2 Existing Parking Systems**

There are three types of existing parking systems. They are 1. Vision Method 2. Sensor Based Method and 3.Two Tier Parking and Automatic Multilevel Car Parking System.

1. **Vision Method**: The detection technology can be divided into two categories. The first category estimates the number of remaining vacant spaces for the entire parking lot by counting incoming and outgoing vehicles. The second monitors the status of each individual space and can be used to guide a car to a vacant space. To detect the status of an individual parking space there are many methods can be used such as ultrasonic waves or surveillance camera placed at different positions.
2. **Sensor Based Method**: Another detection technology uses the sensors to detect the vacant spaces in a parking lot. Different factors play a role in choosing the proper sensor including size, reliability, adaptable to environmental changes, robustness and cost. The sensor technologies can be categorized as either intrusive or nonintrusive. The intrusive sensors can be installed on pavement surface so that digging and tunneling under the road surface are required. Nonintrusive sensor only require fixing on the ceiling or on the ground. Ultrasonic sensors are categorized as nonintrusive sensors.
3. **Two Tier Parking & Automatic Multilevel Car Parking System**: Two tier Parking system is ideally used for people with two cards. The people with two cars can use a single car parking space to park their single car. The system consists of a single platform where the car which is not used often can be parked on the upper level of the platform and the one which is used frequently in the lower platform.

**2.3 Performance Metrics**

In order to access the performance of the strategies implemented in the smart parking systems are 1. Walking Distance and 2. Traffic volume. This also reflects the willingness of the drivers, concerns with the traffic congestion and effects on the environment.

1. **Walking Distance:** The walking distance is defined as distance from the selected parking spot to his destination. This is the parameter also gives a lot of details how happy the drivers are using our parking system. Most of drivers prefer to park closer to their destination. The shorter the walking distance the better it is for the driver.
2. **Traffic Volume:** In our model we have taken traffic model is specifically defined as the amount of traffic generated by search for parking. We cannot negate this factor and is associated with traffic congestion and air pollution.

**2.4 Challenges**

Given our design objectives of our system we require coordination among multiple parties. The main consideration under our system architecture are mentioned as follows:

1. **Fake Parking Requests**: The system collects the data about the performance metrics. The data includes the status of parking space, reservation time, parking location and the driver’s identity. As per the guidelines of the system, the driver can reserve only one parking spot at a time. Sometimes we have multiple parking requests from the same driver. To solve this issue we have implemented the options called Quick Response code. Once the unique identification code is entered into system the information about the driver’s identification, parking location and reservation time. Also to overcome the duplication of parking requests there is a buffer queue which will check the new request and compare it with the existing id requests and help us identifying the fake requests for parking or duplicity of the parking systems.
2. **User Identification Verification**: The user identification is a major security concern as the users with no reservation can enter and occupy someone else’s parking space. The quick response code acts as the unique identification element in the system.
3. **Delay in Parking**: After the reservation of the parking slot the driver will have thirty minutes to reach the parking slot. If in any case the driver is late to reach the spot the timing will be extended. The driver will have to pay for the extended time.
4. **Timer:** As use as the driver reserves the parking slot, the driver will be notified the time till which the driver has the parking slots reserved. Once the driver reservation time is nearing its thirty minutes deadline. This will help the driver ensure that he/she arrives at the parking slot within stipulated time.

**3. Contribution:** We in this paper intend to extend the concept of Smart Parking reservation system to also understand the trends of the various pollution context. We chose the ozone index as the key parameter for our analysis. Ozone layer is one of the important layers in the atmosphere. This helps in keeping the atmosphere cool and protecting the earth’s surface from the ultraviolent rays. The intrusive sensor is installed directly on pavement surface so digging and tunneling under the road surface are required.

4. **Data**: For our analysis we have used the two datasets of city of Aarhus in Denmark. They are 1. Parking Data Set and 2. Pollution Data set. The parking data set consists of the following 1. Vehicle Count 2. Total Space 3.GarageCode. The pollution data set has the following the data points. They are 1. Ozone 2.Sulpur Dioxide 3. Particulate matter 4.Nitrogen Dioxide and Carbon Monoxide. We now look at each of the data points in details.

**4.1 Parking Data Set**

We look at the each data field in the detail below:

1. **Vehicle Count**: This gives the number of vehicles which can be parked in the garage in a streets of Aarhus city in Denmark.
2. **Total Spaces**: This gives the number of the spaces available for parking in a garage in the Aarhus city of Denmark.
3. **Garage Code**: This is the unique identification code for the garage where the vehicles are parked.

**4.2 Pollution Data Set**

The pollution dataset has the following data points for analysis. They are

1. **Ozone**: Ozone as stated above is one of the important components of the atmosphere. The ozone layer forms a protective blanket around the surface of the earth. This protective layer helps in cool down and avoid the ultraviolent rays from reaching the earth atmosphere. With increase in the pollutants will decrease the ozone content in the atmosphere will lead to a phenomenon called global warming.
2. **Sulphur Dioxide**: Sulphur Dioxide is a major air pollutant which is released in the atmosphere by the industries. This also releases when the coal is burnt.
3. **Carbon Monoxide**: Carbon Monoxide is mostly released by the motor vehicles. The Carbon Monoxide when released in the atmosphere will impact the ozone layer.
4. **Particulate Matter**: Particulate Matter is the sum of all solid and liquid particles suspended in the air many of which are hazardous. This complex mixture for instance dust, pattern, smoke and liquid droplets.
5. **Nitrogen Oxide**: Nitrogen oxide are a family of poisonous, highly reactive gases. These are released in the atmosphere when fuel is burnt at high temperature.
6. **Longitude**: The longitude is the geographic coordinate that gives east and west position point on the earth’s surface. This along with the latitude will give the exact position of the place we intend to go.
7. **Latitude**: The latitude is the geographic co-ordinate gives the North and South position point on the earth’s surface. This helps us gets to the accurate position where we want to go.

**5. Method**:

**5.1 System Software Description**

We used the software to make our analysis named Apache Zeppelin. Zeppelin is a web based notebook that enables interactive data analytics. We could make amazing data driven interactive and collaborative documents with SQL, Scala and more. Zeppelin provides built in Apache Spark Integration which will help us build a separate module, library for it. Apache Zeppelin with Spark Integration provides:

* Automatic spark content and SQL context injection.
* Run time jar dependency loading from local file system.
* Cancelling job and displaying its progress.

This software helps us to perform do data visualization on the data. Some of basic charts are already included in package. Zeppelin visualization are not limited to spark sq. query any output from any language back end can be recognized and visualized. Zeppelin aggregates values and displays them in the form of pivot chart with simple drag and drop. We can easily create charts including sum, count, average min and max.

We used the pyspark interpreter in the zeppelin to ensure faster processing and loading of data. This was very useful to ensure that we have faster execution. We later used the R programming interpreter to execute our analysis.

**5.2 Programming Platform**

We used R programming language to run our analysis. R is the most comprehensive statistical analysis package available to use. We can incorporates all of the standardize test model and analyze a comprehensive language for managing and manipulating the data. The specialty of R is that any new technology and ideas are often tried in R. This is one of programming language and environment developed for statistical analysis by practicing statisticians and researchers. The most important feature of R is that it has many graphical capabilities. The validity of software is ensured through openly source unlike closed source software. We used the R interpreter to ensure we have proper analysis.

**5.3 Resilient Distribution Datasets**

As stated above we have used zeppelin to create a Resilient Distribution Dataset (RDDs). The resilient distribution dataset is a data structure in the spark interpret. This is an immutable distributed collection of objects. RDDs can be created through deterministic operations on either data on stable storage or other RDDs. RDD is a fault tolerant collection of elements that can be operated on in parallel. This can used when the data set is huge and cannot be run on a single system. The other main advantage of the RDDs are they are fault tolerance which means that if there is any problem at any node the other segments will take care of it. Creating RDDs have helped our execution of steps.

**5.4 Multiple Linear Regression**

We extended the Smart Parking Information Systems based on reservation to build our system and then used multiple linear regression and time series to help understand the impact on the ozone layer in the atmosphere. Let us understand in the basics of Linear Regression and Time Series Analysis. The 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. In a linear regression explains the relationship between linear predictor function where the unknown estimated parameters. Linear Regression has been useful in many practical uses. Most application fall into board 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 variables to the explanatory values. After developing a fit model more value be given to ensure that model works fine.
* The Linear regression model can also be used to quantify the relationship between the dependent variable and the explanatory variable.

Further 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. The estimation methods of the linear regression models are as follows:

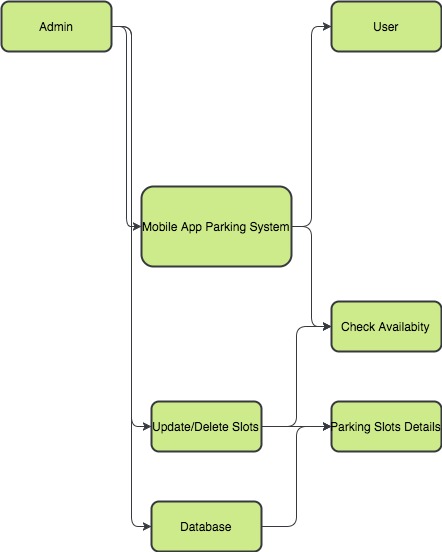
* Ordinary Least Squares: This is the simplest and the most estimator model to be used in prediction of the dependable parameters in the linear regression model.
* Generalized Least Squares: This is an extension of ordinary least squares methods that allows efficient estimation of the beta values different or correlated in many ways.

**5.5 Times Series Analysis**

Later we used the time series to predict the ozone layer in the in the particular in the area of the city. 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 time 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 values based on the historical data we have. A dataset can exhibits both panel data and time series plot. 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 indicates a problem in the analysis or the forecast.

**5.6 System Architecture**

We propose a system which was has all the major parking slots in the city into a database which will be connected to the smart phone of the driver using a mobile application. This database will be regularly updated by the number of vacant spaces available in a particular parking spots. The driver can log in to the mobile application and check for the vacant spots in the different garages across the cities. The driver can select the vacant spot where he/she can park his car. The data given by these systems can used for analysis to understand parking pattern and behavior in various parking slots across the city. This data can be merged with the data opened from the pollution department and perform the analysis on the data and predict the pollution trends in the city.

Now let us look at each other aspect of the system architecture in detail.

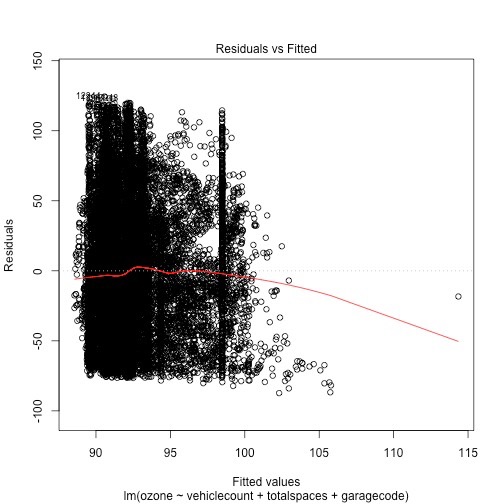
1. **Administrator**: The administrator of the system is responsible for the loading and maintain the number of parking slots in the city. The administrator is responsible for managing the systems for the parking information systems.
2. **User**: The user is the drivers who are in search for parking or request for parking.
3. **Database**: The Database is used to load, update and manage the number of the parking slots in the city.
4. **Mobile Parking Application**: The mobile application will be used to check the number of available parking spots and make reservations for the parking spots.

We have concentrated on the effects of ineffective parking arrangements in a city on the environment. To ensure this we have given equal importance to various components in the pollution data set. As mentioned in the method section we have used ozone as the dependent parameter and vehicle count, total spaces and garage code. Based on the above mentioned parameters we performed the linear regression and time series analysis. The first set of results we tried to understand the basic necessities of the data set to perform the linear regression analysis.

**6. Results:**

**6.1 Residual Results**

The first test we performed is the residual analysis to check if we can perform the linear regression analysis on the data set. We ran the initial analysis between the residuals and the fitted values. Based on the results we can see that the data points in the data sets are randomly dispersed along the horizontal axis. This gave us the freedom to use the linear machine learning algorithms.

****

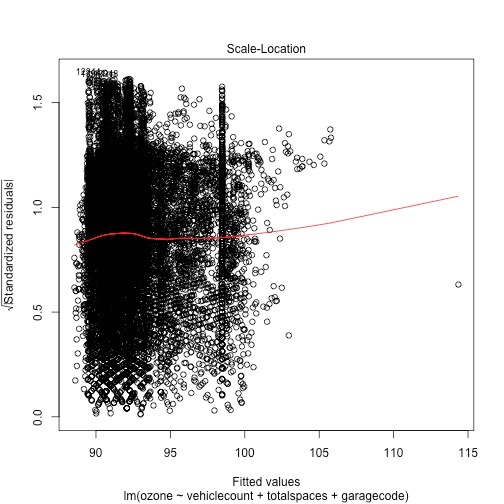
**Figure 6.1 shows us the analysis between the residuals and fitted values**

**6.2 Standard Residuals Analysis**

The next analysis we performed was a standardized residual analysis to understand the distribution of the data points to identify the outliers. This residual analysis help us analyze three major characteristics. They are

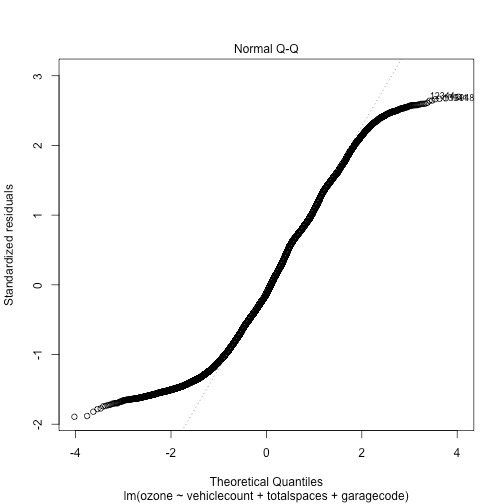
* The residual is bound randomly around the zero line. The suggested assumptions give us the analysis if we can make a linear analysis.
* This also suggests that variance of the error terms are equal.
* It also shows the data points which are outliers.

Based on the results we got we had a clear analysis of the parameters and the data points which are the outliers.



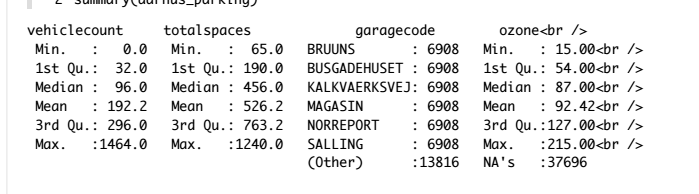
**Figure 6.2 Shows the Graph between Standardized Residuals and Fitted Values**

**6.3 QQ plot**

The QQ plot was the next analysis we performed to identify if any variables are correlated to each other. The points plotted in the QQ plot are always non-decreasing when viewed from left to right. We found the data there were no variables which were correlated which made our analysis easy.

**6.4 Summary Analysis of Linear Regression Analysis**

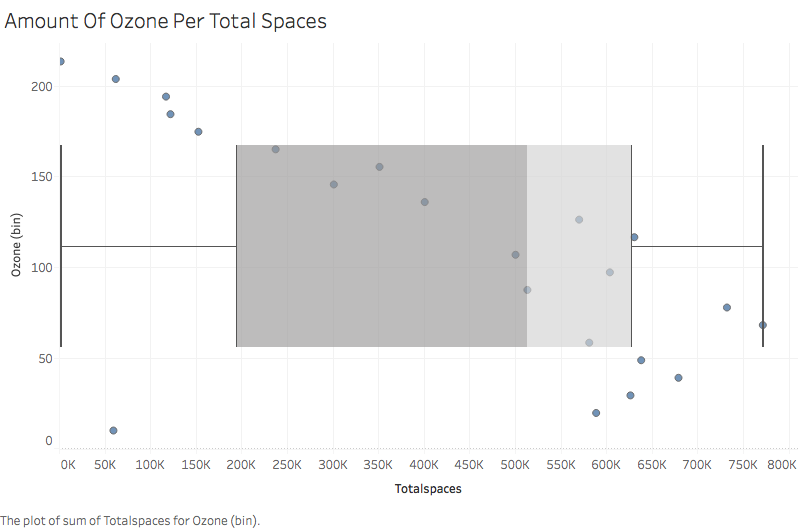
After performing the basic analysis for linear regression model we analyzed the number the summary of the linear regression model to analyze the regression model.



**Figure 6.4 shows the summary of the linear regression analysis**

**Box Plot**

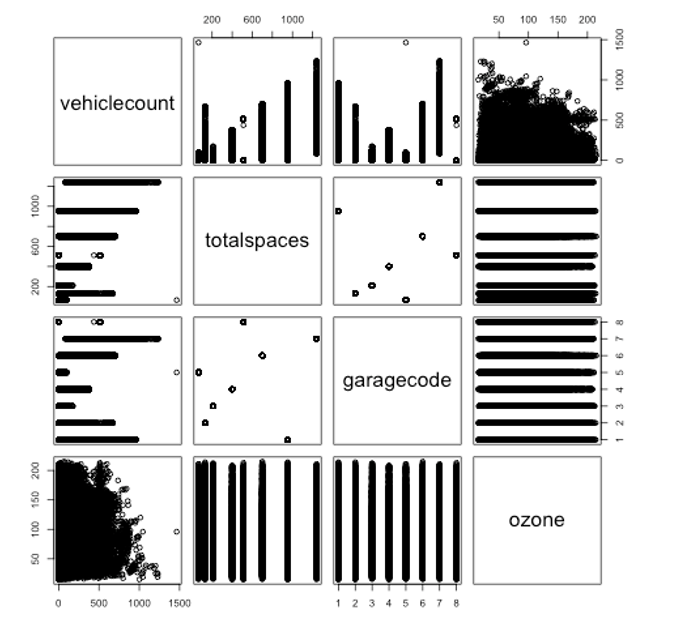
The box plot gives us the information that whether the data is symmetric or skewed. It is a standardized way to display the distribution of the data based on the summary. A symmetric data set shows that the median roughly in the middle of the box. The median is part of the five number summary is shown by the line that cuts through the box in the box plot. Based on 6.4(a) we can say that he has the data is distributed symmetrically.



**Figure 6.4(a) shows the distribution of the total spaces in the data**

**6.5 GG Plot**

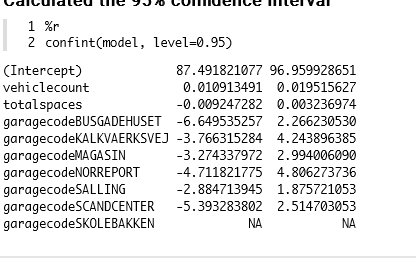
We used the GG plot to get the pictorial analysis of the linear regression analysis. We used the GG plot analysis of the vehicle count, total spaces, garage code and ozone layer.



**Figure 6.5 shows the GG plot of the various parameters**

**6.6 95 Percent Confidence Interval Analysis**

To get the confidence interval of the linear regression we assume that the error e in the linear regression model is independent of x and is normally distributed with zero mean and constant variance. For given value of x the interval estimate for the mean of the dependent variable is called the confidence interval.

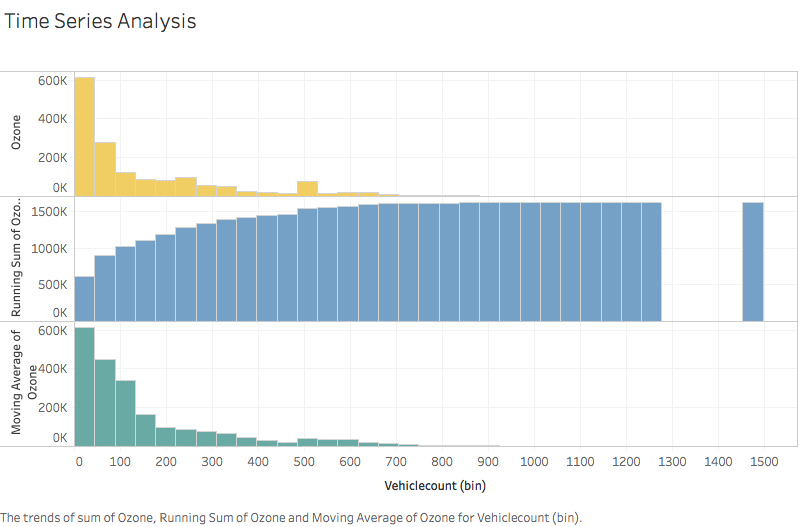


**Figure 6.6 gives the confidence interval of the analysis**

**6.7 Time Series Analysis**

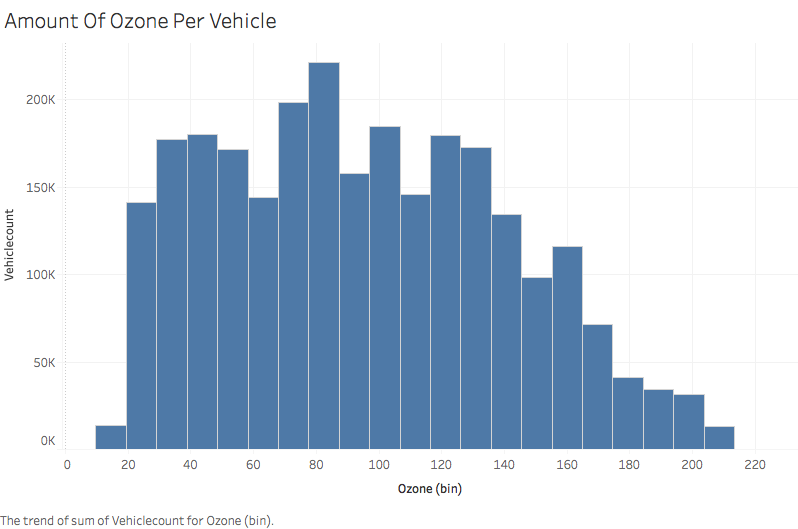
Post the linear regression analysis weperformed the time series analysis to predict and understand the impact of the independent variables on the dependent variables that is the ozone layer. We performed the time series analysis on the ozone layer versus the vehicle count and total spaces in the garage. We can identify where the ozone content is low which indicates the higher pollution and can also help developing better parking infrastructure or better traffic arrangement post tallying the region with the garage code.

We first analyzed the impact of vehicle count on the ozone layer. We found that during busy hours there were areas where during the busy hours as the vehicle count decreases the ozone content in the atmosphere.



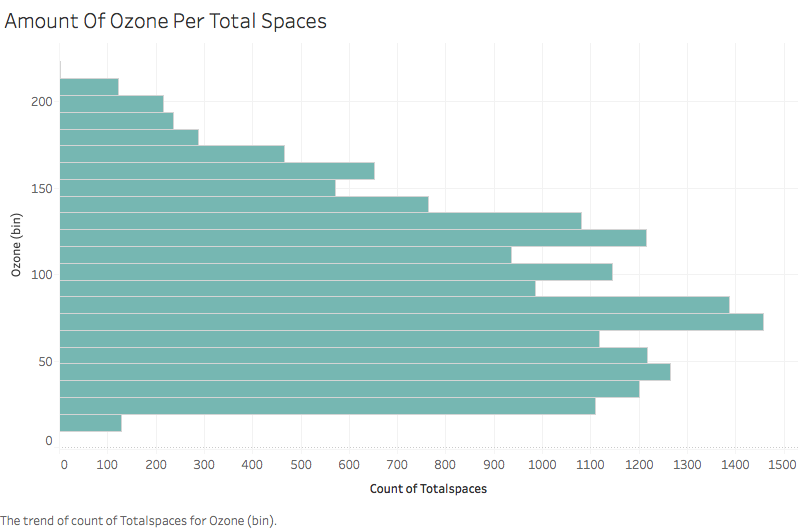
**Figure 6.7(a) gives us the analysis of the vehicle count and ozone layer**

Further we have analyzed the impact on the ozone and the vehicle count in general. As you can see that the increase in the vehicle count and the ozone content in the atmosphere decreases. We identified that during the busy times when vehicle counts are high in a particular region the city which help us support our claim in some area. Hence we check for the amount of ozone layer lost per vehicle.



**Figure 6.7(b) gives us the analysis of the amount of ozone per vehicle count.**

The total spaces also impacted the ozone content in the atmosphere. As the number of total parking spots increases the ozone layer in the atmosphere. This will help us identify the area where the parking spaces are high and design the infrastructure accordingly.



**Figure 6.7(c) shows the impact of the total spaces and ozone**

**7. Discussion**

We took the work of our peers to understand the impact of a proper parking arrangement in major cities will impact the environment. We chose the impact on the environment is this is one matters of concern and consideration across the world. There are a number of a conferences and meetings held across the world to take adequate measures to curtail the air pollution and provide a cleaner environment across the world. The previous work on this topic have concentrated mostly on reducing the amount of time it takes for the driver to find a parking spot in a particular area. The main parameter which was taken into consideration is the parking demand in a particular area. The parking demand is the number of driver who need a parking space in the target area.

We used the real time data of parking data of the Aarhus city in the Denmark. We used the ozone content in the pollution data set. The previous work used the concept of identifying the parking spots at a minimum time. We used the system to collect the data and perform the regression model and the prediction using time series.

The general assumptions made previously made here is that real traffic for parking is proportional to the highway or street traffic. We didn’t have to make any realistic assumptions as we had the garage code which was reference for the target area for the parking and which also helped in understanding the areas where the pollution is high.

The average wait time was the dependent variable which was used in the previous work in the parking management system. In our model we used ozone layer as the dependent variable and vehicle count, garage code and total spaces available in the garage. We used the multiple linear regression to understand the pattern that affects the ozone content in the atmosphere. We later used the time series to predict the levels of ozone layer in the atmosphere. We used the time series to predict the levels of ozone in the atmosphere in a particular target area. This gave us the exact picture of the ozone layer content of the atmosphere with the independent variables. The time series analysis us sixty percent accuracy in the predicting the impact on the ozone layer.

The future scope for this systems will be we can synchronize the mobile application to the Global Positioning system which can help the driver to locate the place where he/she has reserved the parking. The location can be used to also detect the exact location where the pollution is high. This can help the in building better parking infrastructures and making new regulations. Combining the analysis with the global positioning system will help the driver reach the destination within a stipulated time. This will reduce the traffic congestion and also will help reducing the pollution.

**8. Conclusion**

In this paper we have tried to study the impact of the inefficient parking arrangements on the environment. As mentioned previously seventy percent of the air pollution is caused by the search for parking in a city. The carbon monoxide generated will reduce the content of the ozone layer in atmosphere. The decrease in the ozone layer we will lead to the phenomenon called global warming. Hence we chose the ozone content as our dependent variable. From the results we can conclude that the vehicle counts and total spaces available in the garage impact the ozone layer in the atmosphere.

The time series gave us a sixty percent accurate results to predict the areas where the pollution is high. These analysis will be effective in building a smarter city. The department which can benefit from these analysis are 1. Transport department 2. Traffic control department and 3.Infrastructure. The transport department can use these analysis for plan their public transport facilities and their timings which can help reduce the traffic congestion and the also help in reducing the pollution in the environment. The traffic control department can study the pattern of the traffic congestion in an area and also ensure we they can have a proper workforce management in areas where the traffic congestion is reported to be high. And finally the city administration can understand the traffic and parking behavior can help in building proper infrastructure and also regulate the parking prices in the areas where there is a demand in parking during busy hours.

These analysis can be used to build to smart cities. Smart cities are things for the future. Many countries are looking to build smart cities to ensure they can provide a properly connected and transparent administration which can work for the welfare of its citizens.