**Joanna Riascos**

**Dr. Sylvain Jaume**

**Capstone**

**April 28th, 2017**

**Final Teleprompter Script**

**Smart Parking**

For my capstone project, I had to work with the parking dataset. Therefore, I decided to explore if smart parking can help in saving the environment. In doing so, I compared the contribution of the competitor’s article to my contribution. Then I ran a linear regression analysis and a time series analysis. In the sections below, I will discuss further what I did for my capstone project.

**Contribution of Competitor’s Article**

According to the article “A Cloud-Based Smart-Parking System Based on Internet-of-Things Technology” by Thanh Nam Pham, Ming-Fong Tsai, Duc Binh Nyguyen, Chyi-Ren Dow, and Der-Jiunn Deng. The authors have developed a “novel algorithm that increases the efficiency of the current cloud based smart parking system and develops a network architecture based on the Internet-of-Things technology.” The algorithm that they have proposed will help greatly in reducing driver wait time and achieve successful parking. The system will help drivers find a free parking space at the most convenient price. The smart parking system consists of the having a vehicle park as an Internet-of-Things technology (loT) network. It will contain the data that includes the GPS location and distance between car garage areas.

“Smart City” is the city that leads the world. A smart city is a developmental plan vision that implements information and communication technology (ICT) and Internet of things (loT) to manage a city’s assets. It is used in schools, hospitals, community services and power plants etc. The purpose of a smart city is to improve a city’s way of living. It enables mayors and officials to monitor how a city is doing and how the city could improve to provide a better quality of life for its citizens. The cities that have already implemented smart technologies are the following: Amsterdam, Barcelona, Madrid, Southampton, Milton Keynes, Stockholm, and China. Correspondingly, in India they have a “smart cities mission” that was created by the government of India, they have a mission of developing a hundred cities all over the country. The implementation of information and communication technology helps in reducing costs and resource consumption in a city. There are various benefits for having smart technologies employed in cities such as, increased revenue, improved transportation, reduction in crimes, better city planning, recycling, reduction of waste and improved health care. One example is that authorities will have the ability to monitor via the web water and pressure levels. According to a study from the New Jersey Institute of Technology (NJIT), smart technologies “will generate revenues of more than $27.5 billion by 2013.” And 88 cities around the world will adopt these smart technologies by 2025.

Parking is the foundation of a smart city. Smart cities are in the need to look out for reducing parking and street congestion issues. Parking plays a major role in finding a solution to these issues. As reported by the Information Technology Services Reports, “Thirty percent of all traffic congestion in urban areas is caused by drivers looking for a parking space.” Traffic congestion has always been a frightening and global issue that is increasing since there are more drivers in the streets. Also, searching and finding a parking space is sometimes nearly impossible. According to a report, “Smart Parking could result in 2,20,000 gallons of fuels saving till 2030 and approx. 3,00,000 gallons of fuels saved by 2050, if implemented successfully.” With the implementation of smart technologies in parking, the driver should be able to obtain information and see in real-time available parking spaces. As well, it will allow the drivers to reserve their parking spaces ahead of time.

This implementation will reduce car emissions since drivers won’t have the need to drive in circles until they find a parking spot. Similarly, it may reduce stress for the drivers because the majority of times drivers argue over parking spaces.

Many others of the benefits that the cities may have by employing “smart parking” is help traffic in the city to flow more freely, increased mobility, greener cities, optimize parking space usage, enable smart decisions by using big data analytics, monitor and manage available parking spaces, increased revenue, and savings of car fuel. Likewise, there will be savings in paper costs since smart phone applications can take care of the billing aspects of paying for parking reservations. The development for creating “smart parking” consists of building a workflow.

The smart parking workflow analysis consists of six different processes. The first process is the one single gateway covers one area of 1 square Km. Next; the gateway sends the data via the Internet to the database. Then, if a vehicle is parked the occupancy is reported instantly to the users. Afterwards, the central control system gets real-time data about parking occupancy during the day. Subsequently, when a user is connected to the payment system the parking authorities can identify if a driver paid or not. Lastly, when a car is parked the sensor is detected.

**Contribution**

What we believe can help reduce the air pollution produced by the vehicles’ emissions is by implementing smart parking all over the world. An investigation that we have purposed is to explore how smart parking can help solve environmental issues caused by traffic congestion. The purpose is to make cities more sustainable, healthy and happy to live in. By having a smart parking system there will be reduced traffic since there will be less cars looking for an available and open parking space. Also, there will be a reduction in air pollution because it will significantly decrease driving time therefore; it will lower the car emissions.

With the help of smart parking we will be able to contribute in helping save the environment. In addition, with the dataset that we have chosen for our data analysis project we could build models to extract real time parking data and predict future driver behaviors. As an example, a driver might have a daily routine of parking at the same time everyday. If we analyze and extract the data we could have a better understanding of when and at what time the same driver parks his or her vehicle. This will help immensely in having an accurate parking management system to predict when is most likely a parking spot going to be available or vacant. Therefore, as I mentioned in the previous paragraph, I believe we should investigate how smart parking can help solve traffic and environmental issues. We need to find a way to create and implement a smart parking management system in our cities with the help of big data analytics.

**Data Source and Content**

The dataset that we chose to analyze was the parking dataset. The parking dataset comes from the City of Aarhus in Denmark. It contains data from May 22nd 2014 until November 4th 2014 and February 2015 until October 2015. There are a total of eight parking lots. The first Excel csv file contains the following fields: vehicle count, update time, id, total spaces, garage code and stream time. The second csv file has the following fields: garage code, city, postal code, street, house number, latitude and longitude. The data was loaded and read in Zepellin. We imported the apache spark sql function and the date time format packages. Then we adjusted the path and added the location of the data to the script. The data fields that we didn’t include when loading our data was the id, update time, and stream time since we believed that those fields didn’t give us any meaningful information. The vehicle count is the number of vehicles that park in the parking lots, the total spaces is the number of available parking spots, and garage code is just the code to enter the garage. The postal code is just the city’s postal code, the street is where the house is located, the house number is just the house’s number, latitude and the longitude are the geographical coordinates.

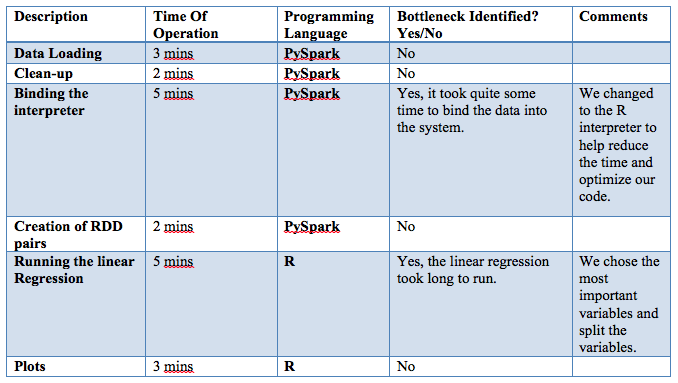
* The parking dataset comes from the City of Aarhus in Denmark. It contains data from February 2015 until October 2015.
* The data file contains the following fields: vehicle count, update time, id, total spaces, and garage code.
* The data was loaded and read in Zepellin.
* We imported the “apache spark sql function” and the date time format packages.
* Then we adjusted the path and added the location of the data to the script.
* The vehicle count field is the number of vehicles that park in the parking lots.
* The total spaces are the number of available parking spots.
* Garage code is just the code to enter the garage.
* Postal code is just the city’s postal code
* Street is where the house is located.
* House number is just the house’s number.
* Latitude and the longitude are the geographical coordinates.
* We merged the parking and the pollution dataset in order to add the "ozone” column.
* The data fields that we didn’t include when loading our data was the id, update time, and stream time since we believed that those fields didn’t give us any meaningful information.

**Method**

The algorithm consists of running a linear regression and a time series analysis. The reason for implementing linear regression is to identify the impact of the parking management system by using different and various tool such as: Zeppelin and Tableau. By running a linear regression we will read in the data and analyze the dataset. We will take as a dependent variable the “Ozone” field and the other fields vehicle count, total spaces, and garage code as the independent variables. What we have done here is that we have merged the two dataset files, the parking and the pollution dataset. We believed that it was a good idea of adding the “Ozone” field from the pollution dataset to our parking dataset since our idea was to investigate how smart parking can help the environment. The reason as to why we chose the “Ozone” column is because car emissions are one of the major contributors of air pollution.

We then ran a time series analysis using the vehicle count and the amount of ozone. First, we took the count of the vehicle count, the sum of the vehicle count, and the rank of the vehicle count. We chose the rank as the data calculation because we wanted to visualize the amount of ozone. We can a time series analysis to determine the impact of the ozone.

**Performance on Big Data: Time Measurements**



For the performance on big data I created a table with five columns, the description, time of operation, programming language, if either a bottleneck was identified or not, and comments.

* For the first row, the data loading took 3 minutes, pyspark was used, no bottleneck was identified, and there are no comments.
* The second row, the cleanup took 2 minutes, pyspark was used, no bottleneck was identified, and there are no comments.
* The third row, binding the interpreter took 5 minutes pyspark was used, a bottleneck was identified (it took a long time to load the data into the system), and we switched the R interpret to help reduce the time and optimize my code.
* The fourth row, the creation of RDD pairs took 2 minutes, pyspark was used, no bottleneck was identified and there are no comments.
* The fifth row, running the linear regression, took 5 minutes, R was used, a bottleneck was identified (the linear regression took long to run) and we chose the most important variables and split the variables.
* The sixth row, the creation of the plots took 3 minutes, R was used, and no bottlenecks were identified.

In conclusion, by running all of these analyses on our model we were able to distinguish which variables had the most impact on our model and which of these variables had the most impact on the ozone layer. In particular, the time series analysis helped us in extracting meaningful data and helped us identify several patterns in the correlated data. This helped us in improving our model. Likewise, with the ending results of both articles we were able to determine that smart parking does help immensely in helping save the environment. Additionally, by implementing smart parking we will be able to spend less money on gas, reduce traffic and stress, improve air quality, real-time data and trend insight, new revenue streams, increased safety, optimized parking, less accidents, less time spent, decreased management costs, and enhanced driver experience. As well, it helps in reducing traffic congestion and reducing the amount of traffic volume while searching for a parking spot.