Literature Survey

Project Title: Shared car locations

Literature Survey:

The massive size of the car sharing dataset and the complexity of geospatial analysis impels us to use big data tooling for our project. BigData queries have become very crucial for the growth of the business. It provides tools for the analysis of large and complex data.

We have categorized the literature survey into the respective domains of technologies used. Papers in ***category 1*** provide insights on the usage of BigData tooling and techniques commonly used for the analysis of large and varied datasets.

Papers in ***category 2*** focus solely on the usage of BigData tooling for spatial analysis of the obtained dataset of car sharing by TelAviv. This section gives us insight on what kind of analysis could be done on such a dataset.

Papers in ***category 3*** describe the complexity of analyzing spatial and temporal data. The main information in this category is about the application of spatial joins with or without the big data environment. Spatial joins are really complex computations and require a lot of in memory processing.

Paper in ***category 4*** shows past reports on an analysis of car-sharing data. There are some very valuable insights which we could use for the analysis of our dataset. We have further modified these analysis approaches to meet our project outcomes like prediction of availability of cars.

The Report Taxonomy can be found in Tabular format at the end of the literature survey report.

Categories:

1. Big data and Distributed computing (only)

Paper 5, generates association rules between items in the large datasets. Estimation, pruning techniques and buffer management is used in the algorithm. This paper is helpful in knowing how to work with large and varied data and applying

association rules.The paper uses Apriori Algorithm(Candidate generation), AIS algorithm, SETM algorithm to do SQL queries on large data items. The performance of each algorithm is plotted on each transaction scale.

Paper 7, focuses on finding the number of cars in restricted urban areas and predicting target time and location based on models generated by the system.Short term forecasting of car availability in restricted urban areas. Machine learning is used for forecasting of multivariate, heterogeneous data like target location, neighboring areas, temporal data, weather, etc.
The data used has features such as past usage patterns, weather conditions, geographical positions and point of interest areas.Regression machine learning models have been applied and error analysis has been done.

In paper 8, emphasis has been provided on data collection using the system they termed as UMAP that is capturing real time data for car sharing websites. The paper aims to understand driving patterns and user habits using data analysis on the data scraped. Data acquisition algorithm has been written to query large datasets.A large json file is downloaded and then different data information is retrieved. Different FFCS algorithms have been used to do analysis on 52 days long datasets, on how users drive in different periods of the day and their driving habits.Different google map API's are used to find available parking spots for such a huge amount of data in real time.

In Paper 13, they talk about the environmental benefits of using a bike sharing system, we analyzed and referred to the Trip distance and average fuel usage estimation that they have implemented so that we can use the similar estimation methodology if required with different estimations for car distance travel and usage of fuel.

2. Big Data in Car Sharing Systems (Our Proposed Project)

In Paper 1, the paper addresses the fundamental concept of the car sharing system i.e. finding out the statistics about locations i.e. depot stations using big data systems. The model further performs predictions using novel deep learning methods with SAE and the first layer of the model as logistic regression to further understand the demand for a location/depot station using the spatial and temporal correlations characteristics of the data. The paper uses Hadoop and Map Reduce to implement the big data system. The paper also compares their results with existing works which implemented RBF, BPNN, and SVM models. This paper was able to achieve better accuracy results.

In Paper 10, the dataset that they used is highly dynamic and non stationary, we referred to the usage patterns of Big Data and how to perform analysis on millions of data references. They have developed a discrete-event trace-driven simulator to study the usage of a hypothetical electric car sharing system. We analyzed the given simulator to see if we can create a similar simulator for making more data references in our dataset if required.

In Paper 11, we learned that In terms of analytical functions of the Spatial Big Data platform, spatial computation through the platform such as the buffering analysis used for the accessibility

analysis of public transportation was found to be efficient in repetitive spatial operations of the massive amount of data. That gave us an idea about what all methods are repetitive and which are efficient.

In Paper 12, The paper was focused on car sharing data in one place like the main idea from our dataset source, from this paper we referred to how to use visual analysis on Big Data. The paper focused on the data of car sharing users and our project focuses on car sharing data of the cars, but we took inspiration from the distribution visualization methods used in the paper.

3. Spatial Joins and Machine Learning (with or without big data)

In Paper 2, the author presents a work with spatial and temporal patterns of car location demand and understanding the usage with respect to spatial data and time of the day (morning, evening, and also further understanding the type of location such as business/residential). Furthermore, the paper presents a classifier to understand the statistics of car-sharing models such as the pickup rate of a car and car availability in a location to understand the profitability of a car station.

In Paper 1, This paper talks about the use of parallel and distributed computing platforms such as Hadoop and map-reduce to tackle the heavy computation issue faced while performing spatial joins to understand the demand for a location/depot station using the spatial and temporal correlations characteristics of the data.

In paper 4, it discusses the importance of car-sharing services, it tells us the effective use of car depots being used for the same and the traffic that it comes across which might hamper the use of the service. It uses the clustering algorithm to group the car of the services and helps predict their better use.

In Paper 14, This paper talks about the importance of car sharing and vehicle sharing in today's world and its importance from an economic point of view. It is concerned about the spatial locations of the vehicles and helps predict them in a better manner to advance the use of the same. It uses the spatial database systems for the same and works on the database of car dealerships and helps it use spatial resources in a better way.

In Paper 16, This paper discusses the importance of Free-Floating Car-Sharing(FFCS) services and how it has boomed in today's world. The car-sharing services need to manage the resources in a very restricted manner to provide great services. The paper helps in achieving that by allowing us to predict the pattern of car-sharing services required in different areas and the resources available in those areas. It helps us predict the number of cars required in the area, space available in the area, and demand for service in that area. It uses a Regression model to make these predictions.

4. Data analysis and reports related to car sharing:

In Paper 18, This paper presents an analysis of bike ride details in combination with calendar events and weather condition datasets to provide new insights into the Spatio-temporal behavior of ride-sharing.

In Paper 19, This paper explores the spatial price discrimination for the ride-sharing networks to balance out the high-demand and low-demand area costs. We will need this to figure out what parking spots are more profitable in comparison to others.

In Paper 6, Studies of urban traffic patterns and traffic congestions. Using carsharing as a means to reduce the congestion of cars in the area.The paper shows analysis on usage and average distance on datasets of five car sharing vendors in italy.It also generates statistical data and generates plots on this scraped data.The analysis helps in finding usage patterns on distances, peak time windows, hotspots, peak days of business, movement trends and frequencies.

In Paper 9, This paper helps us to understand the importance of location or spatial data which helps us to make predictions for our car-sharing systems. The real-world data of car-sharing systems might be incorrect or incomplete and hence can not be used for predictions. This paper helps us to generate car-sharing data for relocation simulation by using machine learning models like the Gaussian Mixture model and two-class classification tree model to generate artificial testing data

In Paper 3, the work is a master's thesis wherein different regression techniques like Support Vector machines, and other techniques such as DecisionStump. Further data mining techniques such as the weka perception model have been deployed and in-depth analysis has been provided. The paper uses feature engineering to combine data features of different slots of data, days of week, and areas of the city. The paper uses data from three different cities, providing analysis of each city separately as well as finding a correlation in driving habits between cities. In Paper 17, is a university report further discussing the factors like in paper 3 using regression models of Decision Tree and aims towards potentially leading to decrease in road congestion for car sharing programs.

References:

1. **Zhu, Xiaolu, et al. "Optimization approach to depot location in car sharing systems with big data."** ***2015 IEEE International Congress on Big Data*****. IEEE, 2015.**
2. **Boldrini, Chiara, Raffaele Bruno, and Marco Conti. "Characterising demand and usage patterns in a large station-based car sharing system."** ***2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*****. IEEE, 2016.**
3. **CAGLIERO, Luca, Paolo GARZA, and Marco Delrio. "Analysis of car sharing data and service availability prediction using data mining techniques."**
4. **Liu, Zhihan, Yi Jia, and Xiaolu Zhu. "Deployment strategy for car-sharing depots by clustering urban traffic big data based on affinity propagation."** ***Scientific Programming*** **2018 (2018).**
5. **Daraio, Elena, et al. "Predicting car availability in free floating car sharing systems: Leveraging machine learning in challenging contexts."** ***Electronics*** **9.8 (2020): 1322.**

6. Trentini, Andrea, and Federico Losacco. "Analyzing carsharing “public”(scraped) data to study urban traffic patterns." *Procedia Environmental Sciences* 37 (2017): 594-603.
7. Cagliero, Luca, et al. "CarPredictor: forecasting the number of free floating car sharing vehicles within restricted urban areas." *2019 IEEE International Congress on Big Data (BigDataCongress)*. IEEE, 2019.
8. Ciociola, Alessandro, et al. "UMAP: Urban mobility analysis platform to harvest car sharing data." *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2017.
9. Brendel, Alfred Benedikt et al. “Generating Rental Data for Car Sharing Relocation Simulations on the Example of Station-Based One-Way Car Sharing.” *HICSS* (2017).
10. Cocca, Michele, et al. "Free floating electric car sharing: A data driven approach for system design." *IEEE Transactions on Intelligent Transportation Systems* 20.12 (2019): 4691-4703.
11. Choi, Junyoug, and Jeakak Yoon. "Utilizing Spatial Big Data platform in evaluating correlations between rental housing car sharing and public transportation." *Spatial Information Research* 25.4 (2017): 555-564.
12. Yu, Daben, and Zongping Li. "Study on Users' Travel Behavior of Urban Car Sharing System Based on Spatio-temporal Big Data in Chengdu." *Journal of Physics: Conference Series*. Vol. 1910. No. 1. IOP Publishing, 2021.
13. Barulli, Michelangelo, et al. "On scalability of electric car sharing in smart cities." *2020 IEEE International Smart Cities Conference (ISC2)*. IEEE, 2020.
14. Lage, M. O., Machado, C. A. S., Berssaneti, F., and Quintanilha, J. A.: A METHOD TO DEFINE THE SPATIAL STATIONS LOCATION IN A CARSHARING SYSTEM IN SÃO PAULO – BRAZIL, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-4/W11, 27–32, https://doi.org/10.5194/isprs-archives-XLII-4-W11-27-2018, 2018.
15. Agrawal, R.; Srikant, R. Fast Algorithms for Mining Association Rules in Large Databases. In Proceedings of the 20th International Conference on Very Large Databases, Yasmine Hammamet, Tunisia, 6–9 October 2002; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA
16. occa, M.; Teixeira, D.; Vassio, L.; Mellia, M.; Almeida, J.M.; Couto da Silva, A.P. On Car-Sharing Usage Prediction with Open Socio-Demographic Data. Electronics 2020, 9, 72. https://doi.org/10.3390/electronics9010072
17. Tsao, H-S. Jacob, and Da-Jie Lin. "Spatial and temporal factors in estimating the potential of ride-sharing for demand reduction." (1999).
18. Corcoran, Jonathan, et al. "Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events." *Journal of Transport Geography* 41 (2014): 292-305.
19. Bimpikis, Kostas, Ozan Candogan, and Daniela Saban. "Spatial pricing in ride-sharing networks." *Operations Research* 67.3 (2019): 744-769.

Report Taxonomy

Paper Number	Paper Type	Category
1	**Research paper**	Big Data in Car Sharing Systems, Spatial Joins. and Machine Learning

2	**Research paper**	Spatial Joins, Data Analysis and reports related to car sharing
3	**Master's Thesis**	Data Analysis and reports related to car sharing
4	**Research paper**	Big Data in Car Sharing Systems, Spatial Joins and Machine Learning
5	**Research paper**	Big data and Distributed computing, Spatial Joins and Machine Learning
6	**Research paper**	Data Analysis and reports related to car sharing
7	**Research paper**	Big data and Distributed computing
8	**Research paper**	Big data and Distributed computing
9	**Research paper**	Data Analysis and reports related to car sharing
10	**Research paper**	Big Data in Car Sharing Systems
11	**Research paper**	Big Data in Car Sharing Systems
12	**Research paper**	Big Data in Car Sharing Systems
13	**Research paper**	Big data and Distributed computing
14	**Research paper**	Spatial Joins and Machine Learning
15	**Research paper**	Spatial Joins and Machine Learning
16	**Research paper**	Spatial Joins and Machine Learning
17	**University Report**	Spatial Joins and Machine Learning
18	**Research paper**	Data Analysis and reports related to car sharing
19	**Research paper**	Data Analysis and reports related to car sharing