How to Cite:
Ramesh, S. S. S., Pawar, K., Dube, R., & Gopan, R. (2022). Analysis of taxi dataset to categorize city rankings. *International Journal of Health Sciences*, *6*(S1), 7625–7638.
https://doi.org/10.53730/ijhs.v6nS1.6649

Analysis of taxi dataset to categorize city rankings

S. S. Subashka Ramesh
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennia-600087, India
Email: subashka@gmail.com

Kshitij Pawar
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennia-600087, India
Email: kp9293@srmist.edu.in

Raghavi Dube
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennia-600087, India
Email: rd2656@srmist.edu.in

Rahul Gopan
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennia-600087, India
Email: rg2994@srmist.edu.in

***Abstract*---**The city administration values the district attraction rating since it can aid in extracting the desirability of the location and hence support the officials in making smart city development decisions. Traditional urban planning tactics mostly rely on Gross Domestic Product, rate of employment, number of people per unit area, and district statistics gleaned through surveys and questionnaires, among other factors. As a point of reference, such knowledge becomes less and less helpful over time. The volume of urban data is growing at an exponential rate. Furthermore, these tactics suffer from a fatal flaw: they are unsuccessful. Independent representations of a district's appeal, as well as inter-district interactions, are not taken into account. It is now feasible to use urban data efficiently for urban planning thanks to advances in urban computing. To that end, this paper proposes PageRank, a district attractiveness rating algorithm based on taxi big data, which is the first to do so. A working software is constructed for visualization motives. To begin, the total area is split into numerous parts using the k-means algorithm. Second, the

Manuscript submitted: 18 Feb 2022, Manuscript revised: 09 March 2022, Accepted for publication: 27 April 2022

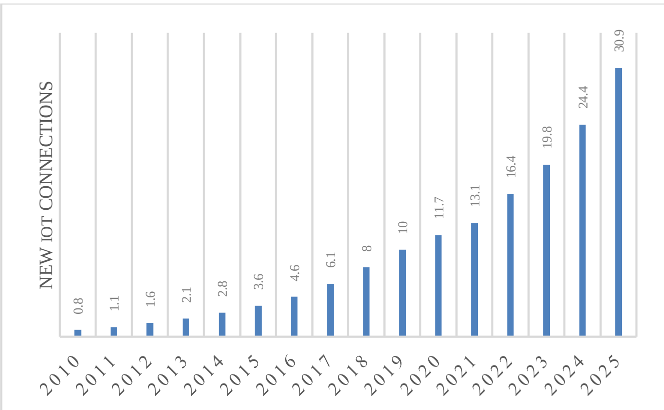
pagerank algorithm is enlarged to incorporate the journey's Origin point Destination point in order to generate the Origin and Destination matrix, which may be used to calculate the district's attractiveness rating. Eventually, by exhibiting the findings generated by PageRank, we can effectively extract the pattern of how district enticement change as time passes and fascinating discoveries on city life. As a result, it has a wide range of applications in urban planning and data mining.

***Keywords*---**Index terms – District Rankings, Taxi Data Set, K-Means, O-D Matrix, Big Data.

Introduction

Traditional municipal implementation methods depends mostly on generous metrics such as district Gross Domestic Product, rate of employment, people in an area and data from survey questions, those can be considered as not efficient means of planning and hence yield unsatisfactory results. Furthermore, mentioned procedures or rules do not prove to be effective as compared to when big data was booming. With the rapid advancement of precise locating technology, obtaining a big number of travel data is becoming accessible, and determining trip appealing regions where transportation and travelling are recorded in a higher numbers tends towards new area of study emphasis.

When information and communication technologies (ICT) are used to improve urban infrastructure, a Smart City arises [1]. The Internet of Things (IoT) paradigm [2] has spawned a slew of new communication protocols that may be used to communicate with commercial products utilizing a variety of information visualization. Considering the case, an automation based programme is required to control the elements of interoperability and to consider the incorporation of satisfactory Artificial Intelligence (AI) algorithms for modelling contextual relationships.



Graph 1: Number of IoT devices to be installed in upcoming years

Table 1
Sample Information from Taxi Data Set

Information	Vehicle Record 1	Vehicle Record 2	Vehicle Record 3	Vehicle Record 4
Vehicle ID	VID 1	VID 2	VID 3	VID 4
Vehicle Number	1	2	3	4
Weight	157739.283099506	86429.3481587842	56925.2638757417	66829.0382771858
Origin	113.323998 23.0992071	113.358151 23.098927	113.377737 23.094384	113.340401 23.078534
Destination	113.314744 23.098706	113.349660 23.089066	113.364049 23.085386	113.331468 23.088039

There is a plethora of data sources available in metropolitan settings. A large number of sensors are scattered across cities, the majority of them are put in indoor locations. This circumstance has resulted in the development of new investigative methods and technologies that put light on cases, assisting to manage the machines in a more efficient and collaborative manner [3]. There are also a variety of movable information sources, such as mobiles, smart watch, sensing gadgets, and on-board capture devices in cars. All of capturing devices offer data that allows for the detection of city journey patterns. Almost every contemporary city administrating devices are unable to efficiently utilize this huge data available, resulting in large volumes of data being underutilized. Many AI approaches in Computer Science have been proposed in this field to cope with filtering of massive amounts of information in order to analects meaningful details [4], respectively stated as big data.

Literature survey and existing system

The impact of several factors on the overall fee amount paid to customers is investigated and studied in this study, which uses taxi trip data from 2015 to give insights for consumers in New York City to plan their travels in the most cost-effective manner. [1]. Geospatial information which can be extracted from locating devices which are used in numerous instruments used in our day to day lives are important to be ready or prepared in case of geographical emergencies such as earthquakes etc in order to evacuate or take proper action during the time of emergencies and several other cases where we have to alleviate of the respective problems risen. The introduction of a worldwide navigation system and non-wired communication technologies revolutionized how we live and acquire geospatial data in the field. [2]. A taxi stand can efficiently manage taxi driver conduct, minimize empty-fare rates, and offer a comfortable and systematic holdback atmosphere for the general public. Nevertheless, most cities' existing taxi stands are set up in an inefficient manner, resulting in a poor usage rate and a waste of public space resources. This study proposes a unique three-stage technique for progressively addressing the taxi stand location issue (TSLP). [3]

The municipal government values the district attraction rating since it may be used to reveal the district's attractiveness and enhance government decision-making in terms of urban development. Traditional urban planning tactics rely on a variety of indicators, including district Gross Domestic Price, rate of employment, people living in an area, and data from surveys. In contrast, as the volume of urban data expands, such data becomes increasingly useless. Furthermore, these methods have a basic weakness in that they are individual representations of a district's attractiveness that ignore the interplay across districts. It is now feasible to use city information efficiently for city management thanks to advances in urban computing. To that end, this research offers PageRank, a district attractiveness rating algorithm based on taxi big data collected which is the cab data set has been used for city attraction counting. A working model is constructed for visualization needs. To begin, the given area is partitioned to form small parts so as to easily use for calculation purpose in different algorithms efficiently.

Second, given approach is enlarged to incorporate cab Origin point and Destination point in order to generate the Origin and Destination graph, which may be used to establish the district attraction rating. Finally, we can effectively acquire the pattern of how district attractions vary over time and intriguing discoveries on urban by showing the results and case studies that have been developed. As a result, it has a wide range of implementation in city management and smart city generation. The working model is now available for use. PageRank, a new technique for measuring district attractiveness using taxi operating data, is used for the first time in this study. PageRank's attraction index, in contrast to other district indices such as GDP and district population, may describe the interaction between such districts.

Output 1: Dataset Sample

```
In [8]:  1  print(df.head(5))

   id  from_idx  to_idx  date_month  date_day  date_hour  route_weight
0   0         0       0           2         1          0             9
1   1         0       1           2         1          0             1
2   2         0       2           2         1          0             2
3   3         0       3           2         1          0             1
4   4         0       4           2         1          0             3
```

Proposed system

According to recent research the recommended strategy presupposes that the majority of customers plan to share a taxi ride. Passengers may either convey their trip demands directly to the model using management system or indirectly by hailing a shared cab on the side of the road. The system will identify appropriate cabs to answer both insystem and ofsystem journeys by calculating the availability of all cabs and creating routes for the chosen taxis depending on specific service demands. The proposed system specifies two main procedures: request handling and assignment, as well as remediation approaches for coping with unanticipated journey problems. Within the research, the chance that certain unforeseen road circumstances may impact the time up of a group of journey plans given a specific trip schedule assignment might alternatively be considered as a driver's trip planning task. In the suggested technique, just the origin Riding Fare of a trip request is used to discover potential taxis. Despite the fact that both the origin and the destination are employed in a dual-side search, these works consider them separately. Furthermore, rather than searching for the greatest taxi, previous work will be resumed as soon as a legitimate taxi is found. Passenger-taxi matching based on insufficient trip information is unable to identify and exclude invalid taxis. While it is possible that it will miss the best taxi that introduces the lowest cost at first, it is also possible that it will not be found to be the cheapest amongst the others, the lowest cost afterwards. in this collection of work

Data preprocessing

1. K-Means Clustering
K-means is a clustering approach that generates a predetermined number of clusters from point data. Minor way K-means is a simplified way of the algorithm that is used to cluster a large number of points. Constraints those cannot be linked are known as constrained K-means [5]. It is necessary to give cannot-link constraints, which is difficult to provide in practice. Another way we can tackle this problem is by collecting numerous polygon ends or linings those points from distinct polygons are unable to connect to one another. Then, in each polygon border, Mini Batch K-means is used to create clusters [28].

2. Contour Extraction
Contour extraction is a method for extracting contours from point data. Convex Hull and Alpha Shape are two approaches that can be used for distinct objectives. Convex Hull every time gives the same output in the form of polygons which are sophisticated that only fits the points, whereas Alpha Shape creates polygons but those are different or opposite of convex known as concave, depending on the setting settings, may keep the shape of the given points in some situations. We chose the Alpha Shape as the contour extraction technique to benefit a more attractive visual that exaggerates appropriate for the plotting. Otherwise, Convex Hull is sufficient for extracting point contours. To explore extreme weather events utilizing advanced way of computing which is not physical [6]. Chang described the data processing architecture and ideas of P-Map, PMerge, and P-Reduce, as well as a greedy algorithm for completing huge information procedure along with taking available end points into account [7].

3. District Generation and Analysis
While conventional district creation depends upon the performance of the managing cities, where certain cities are frequently of unusual forms and are associated with higher commandment, it is not appropriate in our work's attractiveness analysis. This is due to the fact that within a single administrative district, there may be multiple sub-districts with varying attractions. To do this, the clustering technique is used to create districts by grouping almost numerous locating points generated from start and end of the location gathered over the course of 8 weeks. Due to the complicated geographical distribution features of congested roads transportation data, the K-means technique is used to automatically build districts by grouping all of the origins and destinations coordinates.

As a result, Constrained K-means [5] is used to account for with polygon bounds picked non automatically making use of an online tool called map2tool. Considering the congestion number estimation approach proposed in [8], it appears to be appropriate to divide city into number of areas, each of which is a city. Nevertheless, in this example, because certain clusters are situated in locations far from metropolitan areas, we must manually add some central locations to these areas to make sure that these locations have adequate area centers, which totals to X centers. By deleting a city that is located on an contactless area with no cab data, a total of Y districts are obtained. We perform calculations of each city taking reference of the start and end points which sums over the duration of 8 weeks to make the best interpretations of the attractions of several areas in the city and to prepare for further computation. Then we construct the Start End matrix, one of which elements contains the number of data for a cab changing from one district to another with an hour's precision.

Data Analysis

To examine the attractions in various locations, we must first divide the city into several areas whereas old city management was based upon the managing top tier cities which can be seen that some cities are frequently of unusual description and are associated with managing directives, it is not appropriate in our work's attractiveness analysis. This is due to the fact that within a single administrative district, there may be multiple sub-districts with varying attractions. To do this, the clustering technique is used to create districts by grouping almost several junction slots of start and ends gathered over the course of 8 weeks.

Due to presence of sophisticated geographical properties of traffic transportation information, the given algorithm is used to automatically build districts by clustering all of the origins and destinations coordinates. There are several properties making up the connection, organically dividing the area into many regions. As a result, Constrained K-means [5] is used to account for river barriers, with polygon bounds consisting of rivers picked manually using an online map2 tool2.

According to estimation approach described in [8], it appears to be appropriate to divide an average city into several areas, each of which is a city. Nevertheless, in this example, because not all but areas are situated in locations away from metropolitan cities, practitioner must manually increase number of locations to areas to make sure that mentioned non rural area have enough area positionings, which totals X centers, adding up to Y districts are achieved by deleting a area that is located on an contactless area with no taxi data.

We perform computation in every area referencing the start and end information which totals up to over 8 weeks in order to have a better understanding of the attractions of various areas [29]. Then we construct the Origin point and Destination point graph, one of which elements contains the number of information for a cab travelling to another location making use of the data in an hour's accurate duration. Later, a more complete description will be provided.

Modules

1. Microscopic Mobility

Both macroscopic and microscopic aspects are crucial and must be taken into account when creating realistic mobility models. Inaccurate depiction of the mobility model's mobility characteristics, whether macroscopic or microscopic, can lead to inaccurate results and conclusions. Mobility models' microscopic characteristics define and depict the Vehicle behavior may be predicted more precisely. To model the microscopic movement of automobiles, we use an open-source discrete-time traffic simulator dubbed simulation of urban mobility (SUMO). It has the capacity to more correctly represent the vehicle's specific behavior. It also allows you to extract location details and a variety of sources, including new area which are yet to be travelled. It also has portable libraries that provide precise, efficient as well as traffic redirect ability modelling. It offers adaptable solutions for a wide range of research areas, including precise locating as well as intervehicle contact, and navigation, to name a few. It may be used to create travelling graphs making use of large are scopic and microscopic properties, and the generated tracking files can be loaded into network performance trackers.

2. Trip Generation and Distribution

The usage of precise travel models has grown in recent years, as the research community has become more interested in wide applications of public answering for a variety of applications. The scientific community has concentrated on both the microscopic and macroscopic components of vehicle movement. Certain difficulties, such as traffic congestion, have prompted researchers to dig deeper and develop wide applications of public answering for a variety of applications situations, or VSNs. The O-D model is a fundamental component of the vehicular mobility model. The literature's simplistic or unrealistic techniques motivate researchers to develop travelling projects those can be implemented in real life and comparable to relatable conditions. The generation procedure for the described start and end model and constant representation of junctions over time based on trackers are presented in this part. The S-E graph is a representation of a vehicle's mobility pattern in terms of travels from one point to another.

To build a precise Start- End graph, both trip production and trip distribution are equally important. The first step toward a precise tracking system is trip generation. Some of the elements that influence the trip creation process include congestion management restrictions, classification data. It is not possible to model trip generation using a basic technique. The process of trip production needs as much realization as feasible. As a result, we generate accurate data of visits starting from appealing location within a given time interval using the computational references offered in the preceding section. For trip creation, we apply several probability distributions functions with estimated parameter values. The second stage is to generate journey distribution calculated over number of cabs travelled in specific duration that originate in a certain location I and end in a specific area j. To put it another way, we must determine the number of points that link various areas.

The representation for a journey graph, as previously stated, the entire area is split into Z sublocations, and the edges connecting these vertices are represented by the journey traveled in every particular area. Indegree and outdegree describe the attractiveness and not appealing nature of these points, whereas the exact points connecting ends denotes the journeys in total started from a certain place. Of road sightseeing procedures, road photography, and car tracking are some of the methods used to estimate the O-D matrix. We are, on the other hand, concentrated in calculating the S-E graph according to the generated data.

Furthermore, certain approaches for calculating S-E graphs need extra data. One of the strategies used in transportation engineering to create the S-E graph based on traffic counts is the gravity model. The gravitational law of Newton is comparable to this idea. The number of trips created from an origin to a destination is supposed to be directly proportionate to the total number of journeys originating and beginning from a certain place.
The given formula is used to estimate the number of journeys beginning from area i to area j:

$$T_{ij} = \propto \frac{P_i E_j}{{d_{ij}}^2}$$

Where Pi is the population in area i, Ej is the employment in area j, and dij is the distance or travel time between i and j. α is the parameter of calibration.

$$V_a = \sum_{ij} T_{ij} P_{ij} a$$

$$P_{ij} a = \begin{cases} 1 \\ 0 \end{cases} \text{if } T_{ij}$$

3. Traffic Assignment
The quickest and most direct road from the origin to the end location is computed using traffic assignment algorithms. The volume of traffic changes over time. Furthermore, some unique constraints imposed by traffic control agencies on some roads have a substantial influence on the route selection process. For numerous reasons, public vehicles without specific rights aren't permitted to make journey byroad. Similarly, some routes, such as instant travel busses, are designated for specific vehicles to move along them. As a result, choosing the quickest route from point A to point B without previous knowledge of road topology or restrictions imposed by management leads to the adequate but nonrealistic traffic management. The amount of trips on a certain point origin I to end area j has an impact on the traffic assignment process. Calculating the best route from point A to point B is a difficult undertaking, and several considerations impact this process, which includes travelling congestion.

4. Network Topology
The street plan of city from the database is used in our simulation. The map includes information on road signals, topography, and other sites of interest, such as management area. When compared to actual road topology, however, this imported plan reveals inconsistencies in information that influence vehicular mobility. We execute numerous procedures on this imported map in order to create the road network topology equivalent to real-world road networkings, which we discuss in detail in the next section.

Algorithm Execution Comparison

1. Pagerank and Extensions

PageRank is a classic tool for assessing the web's linking structure [9]. It may be thought of as a method for mimicking the patterns of users over a program of online pages linked by hyperlinks [10]. Many efforts have been undertaken to provide a thorough understanding of PageRank theory [11], [12], [13].
The formula is as follows:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

where u, v are website pages, P R(u) is the meaning of website page u, Bu is a set of all the webpages that are backlinks (e.g., an is a backlink of b indicates that a connects to b) of webpage u, and L(v) is the number of external links that originate from webpage v [14]. However, there is a fatal flaw in that dead-ends (i.e., internet pages with no external connections) cause PR values to converge to zero when PR values are calculated using for the black-hole-like influence of dead ends. To solve the above problem an improvised formula has been provided:

$$PR(u) = \frac{1-d}{N} + d \sum_{v = \in B_u} \frac{PR(v)}{L(v)}, s.t. 0 < d < 1$$

The PageRank algorithm may be thought of as a method for simulating the movement of internet users around a network of web pages linked together [15]. The distribution of internet users among online sites is shown by analyzing the tracks of internet users likelihood of redirecting to other PageRank assures the probability of movement between nodes to assess the movement of random surfers. At the start, all nodes

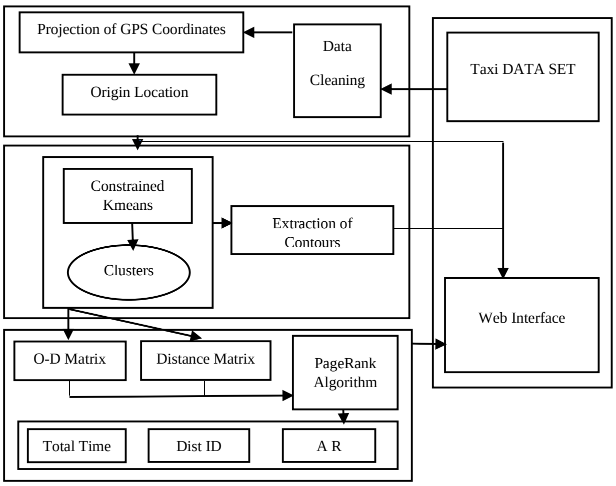
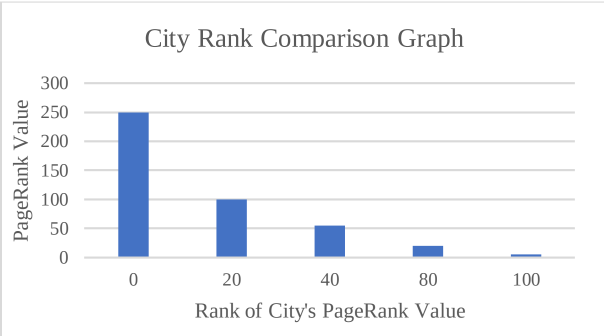


Figure1: Architecture Diagram

are given the same mark. This is further passed across points using repeated procedure inside the network's link structure. PageRank measures the probability of movement between nodes to assess the movement of random surfers. In the beginning, all points are given the same mark. The mark is then passed across points using procedure inside the network's link structure. The PR mark reaches an equilibrium value after a significant number of repetitions in the movement. The probability of ranks that follow connection, as indicated by the link topology, is one of the major methods. The percentage of marks that move on each profile is computed by dividing the source node's mark by the number of out-links. That instance, if a point has some outputs, the mark for calculation of each profile equals 0.33 of the point's mark. While the nodes reached equilibrium marks, the mark for movement on every connection would likewise become equilibrium. A surfer may quit viewing a website and start visiting a different website, considered to be a movement behavior not interrelated, in addition to moving along the network connection. People, on the other hand, travel physically between areas in a geographic network. As a result, migration outside of the network is unlikely to occur in a geospatial network research [16]. To put it another way, we may suppose that everyone moves along the network links when calculating the PR algorithm.

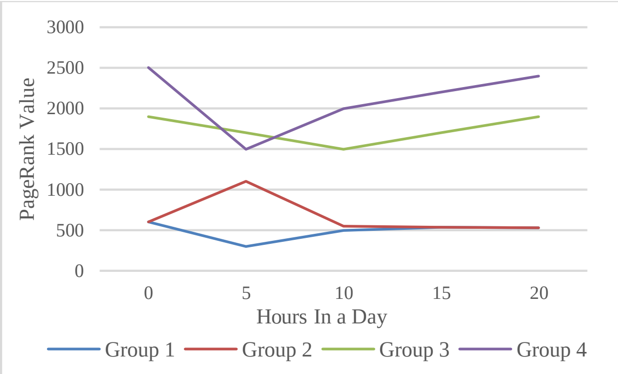


Graph 2: City Rank Comparison Graph

2. Distance Decay Pagerank

The cost of travel is an important geographic issue when approaching a region. Because every terrestrial place contains a defined location, reaching an end from a location of start would have a commensurate trip cost. As travel costs rise, the strength of spatial interconnections between sites drops. The distance-decay effect [17] [18] is what causes this. In a geospatial network, one of the types of trip cost is the geographic distance between nodes. It might considerate problem within connections or an problem affecting the points, with the positions of both nodes determining the outcome.

The PageRank algorithm, especially the Distance-Decay PageRank method, now includes physical distance amongst the points would be as journey cost. We employed the inverse locating amongst the final points and the formula to capture the location degrade effect, and conveyed the PR mark proportionate to the inverse distance from the start to its end points. The size of the location degrade effect is determined by the location factor. The steeper its curve even more exaggerating substantial the its effect will be as the locating factor increases. A locating factor of 0 means that its effect isn't present, and the outcome is comparable to that of the algorithm (2). A sensitivity study was performed to determine the influence of the distance factor on model performance. Furthermore, several research [19] [20] [21] [22] recommended using power operations to simulate its connection. Connecting the calculated results.

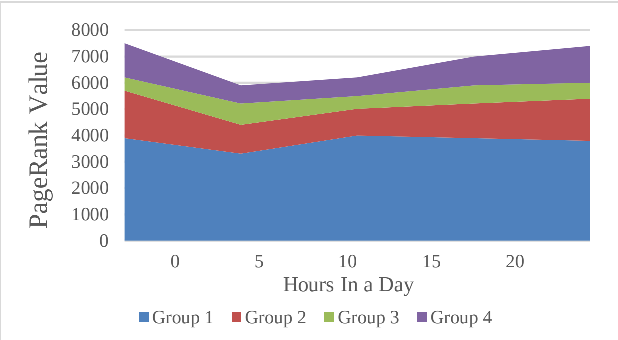


Graph 3: PageRank value comparison with hours (i)

3. Geographic Pagerank

Attractiveness and impedance among locales are common geographic factors. It is expected that the strength of spatial interactions between sites is related to their attraction [23] [24]. Attractiveness in geography refers to the force that draws individuals to a certain location, whereas impedance based on the travelling obstacle that must be overcome [25]. Attractiveness and impedance influence people's mobility across nodes in a geographic network, and these effects may be recorded by computing the likelihood of travel through each connection. We present the Geographical PageRank (GPR) method, which incorporates the ideas of attraction into the PageRank algorithm. We suppose that the chance of individuals choosing closer and more appealing sites is increased, similar to the notion of gravity modelling. In other words, people may walk about nearby linked nodes, but they may also be prepared to go to further away but more appealing nodes. The inbound connections, according to Xing and Ghorbani [26], might indicate each node's direct popularity, which was employed as the comparison weights in WPR [27]. As a result, in this, we define communication of points which reflects attraction, and its travel time from its start point denoting the amount of impedance applied. The target node's mark, FG, is then used to compare it to its comparing nodes (Ob). As a result, a node with a greater mark will be routed to a node that is closer and/or more appealing. Since GPR adds the unique properties functions, the attractiveness of the end node and travel time from start are considered into this calculation are both taken into account at the same time. where t(a) is the mark of the end node an at procedure t, inside coverage (j) is the target node j's in-degree, and is the in-exponent. degree's The distance between nodes I and j are distance (i, j).

In Graph 3 we can see that a comparison between several groups are provided from which the first group is district with most residents, 2nd group can be seen as that of district with transportation hub, 3rd district consists and the 4th district is with entertainment as well as business. Similarly in Graph 4. We can see that the 1st district mainly consists of offices, 2nd district comprises of business as well as smaller components, the 4th district is mainly of restaurants and hotels. In summary, both the DDPR and GPR suggested algorithms include four components. (1) Initialization: every point present in geographic circulation are given an mark at the start; (2) mark process: present for every point, a mark is sent (b), The process for location degrade is calculate and further transferred is calculated using the; then its mark is split and delivered to each of its target nodes, according to the percentage. Calculated based on provided auxiliaries. (3) the process for getting marks is conducted during the procedure of the previous calculations every previous nodes X gets connected to every node Y and hence a mark is generated in order to satisfy and thus the whole practice comes to still and can be further used to proceed with final PR mark generated.



Graph 4: PageRank comparison with hours graph (ii)

Conclusion

In this study, we offer a unique approach called PageRank for estimating district attractiveness using taxi operational data for the first time. The imperial score provided by PageRank can show the connections amongst such locations, as opposed to other district indices such as GDP and population. After then, an app is created to calculate the district's attractions.

- First, the Constrained K-means clustering technique is used to create X districts.
- By calculating the district's rank value for each time depending on the small areas created and cab data set.
- Responsive clustering is used to investigate the personalities of these areas based on the time series of their marks. We can properly estimate the movement of people and further deduce the function of each zone by calculating the outcomes of various areas, which is an essential aspect of urban computing.
- The local city, the transit hub, the management area, the official area, and the combination of the management area and the official are the five types of districts. Every day, taxis pass through their neighborhood.
- There are four significant locating points that are realized. Which could be train or airline leaving time, the cab ride to work time, the after-hours leisure time, and the time when people return home.

Our technology appears to have results in city management and smart city planning, based on our comprehensive research.

References

[1] N. Komninos, “Intelligent cities: variable geometries of spatial intelligence,” Intelligent Buildings International, vol. 3, no. 3, pp. 172–188, 2011.

[2] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” Computer networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[3] L. Da Xu, W. He, and S. Li, “Internet of things in industries: a survey,”Industrial Informatics, IEEE Transactions on, vol. 10, no. 4, pp. 2233–2243, 2014.

[4] Ramesh SS, Rathore KS, Raj R, Vatsalya K, Vatsa M. Integrated Malware Analysis Using Markov Based Model in Machine Learning.

[5] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrodl, “Constrained k-means ¨clustering with background knowledge,” in Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001),Williams College, Williamstown, MA, USA, June 28 - July 1, 2001, 2001, pp. 577–584.

[6] G. Sun, V. Chang, G. Yang, and D. Liao, “The cost-efficient deployment of replica servers in virtual content distribution networks for data fusion,” Information Sciences, vol. 432, pp. 495–515, 2018.

[7] V. Chang, “Towards data analysis for weather cloud computing,” Knowl. Based Syst., vol. 127, pp. 29–45, 2017.

[8] Subashka Ramesh, S.S., Hassan, N., Khandelwal, A., Kaustoob, R., Gupta, S. Analytics and machine learning approaches to generate insights for different sports International Journal of Recent Technology and Engineering, 2019, 7(6), pp. 1612–1617

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web.” Stanford InfoLab, Tech. Rep., 1999.

[10] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” Computer networks and ISDN systems, vol. 30, no. 1-7, pp. 107–117, 1998.

[11] M. Bianchini, M. Gori, and F. Scarselli, “Inside pagerank,” ACM Transactions on Internet Technology (TOIT), vol. 5, no. 1, pp. 92–128, 2005.

[12] P. Berkhin, “A survey on pagerank computing,” Internet Mathematics, vol. 2, no. 1, pp. 73–120, 2005.

[13] A. N. Langville and C. D. Meyer, “Deeper inside pagerank,” Internet Mathematics, vol. 1, no. 3, pp. 335–380, 2004.

[14] A. Rajaraman and J. D. Ullman, Mining of massive datasets. Cambridge University Press, 2011. [15] Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 1998; 30: 107–117. doi: 10.1016/S0169-7552(98)00110-X

[16] Balamurugan, K., Uthayakumar, M., Sankar, S., Hareesh, U. S., & Warrier, K. G. K. (2018). Modeling and surface texturing on surface roughness in machining LaPO4–Y2O3 composite. Materials and Manufacturing Processes, 33(4), 405-413.

[17] Barthélemy M. Spatial networks. Physics Reports 2011; 499(1): 1–101.

[18] Fotheringham S. Spatial structure and distance-decay parameters. Annals of the Association of American Geographers 1981; 71(3): 425–436.

[19] Latchoumi, T. P., Swathi, R., Vidyasri, P., & Balamurugan, K. (2022, March). Develop New Algorithm To Improve Safety On WMSN In Health Disease Monitoring. In 2022 International Mobile and Embedded Technology Conference (MECON) (pp. 357-362). IEEE.
[20] K, Balamurugan, M, Uthayakumar, S, Sankar, US, Hareesh, & KGK, Warrier. Predicting correlations in abrasive waterjet cutting parameters of Lanthanum phosphate/Yttria composite by response surface methodology. Measurement, 131, 309-318, 2019
[21] Latchoumi, T. P., & Parthiban, L. (2021). Quasi oppositional dragonfly algorithm for load balancing in a cloud computing environment. Wireless Personal Communications, 1-18.
[22] Liu H, Chen YH, Lih JS. Crossover from exponential to power-law scaling for human mobility pattern in urban, suburban and rural areas. European Physical Journal 2015; 88: 117. doi: 10.1140/epjb/e2015-60232-1
[23] Karnan, B., Kuppusamy, A., Latchoumi, T. P., Banerjee, A., Sinha, A., Biswas, A., & Subramanian, A. K. (2022). Multi-response Optimization of Turning Parameters for Cryogenically Treated and Tempered WC–Co Inserts. Journal of The Institution of Engineers (India): Series D, 1-12.
[24] Balamurugan, K., Uthayakumar, M., Sankar, S., Hareesh, U. S., & Warrier, K. G. K. (2017). Mathematical modelling on multiple variables in machining LaPO4/Y2O3 composite by abrasive waterjet. International Journal of Machining and Machinability of Materials, 19(5), 426-439.
[25] Rodrigue JP, Comtois C, Slack B. The Geography of Transport Systems 2013, New York, US: Routledge.
[26] Xing W, Ghorbani A. Weighted PageRank Algorithm. Paper presented at the 2nd Annual Conference on Communication Networks and Services Research 2004, NB, Canada.
[27] Jiang B. Ranking spaces for predicting human movement in an urban environment. International Journal of Geographical Information Science 2009; 23(7): 823–837. doi: 10.1080/13658810802022822
[28] Nankani, H., Gupta, S., Singh, S., & Subashka Ramesh, S. S. (2019). Detection analysis of various types of cancer by logistic regression using machine learning. International Journal of Engineering and Advanced Technology,9(1),99–104.
[29] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: Computational intelligence techniques and application areas," Int. J. Inf. Manage, pp. 10–15, 2016.
[30] C. M. Bishop, Pattern Recognition and Machine Learning, M. Jordan, J. Kleinberg, and B. Scholkopf, Eds. Springer, 2006.
[31] Balamurugan, K., Uthayakumar, M., Sankar, S., Hareesh, U. S., & Warrier, K. G. K. (2018). Effect of abrasive waterjet machining on LaPO4/Y2O3 ceramic matrix composite. Journal of the Australian Ceramic Society, 54(2), 205-214.
[32] Banu, J. F., Muneeshwari, P., Raja, K., Suresh, S., Latchoumi, T. P., & Deepan, S. (2022, January). Ontology-Based Image Retrieval by Utilizing Model Annotations and Content. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 300-305). IEEE.