GEO BASED CUSTOMER INSIGHT 360⁰ APPROACH

PRAVEEN KUMAR[1], PROF. (DR.) MOHIT GANGWAR[2]

[1]M.Tech. Research Scholar, Department of CSE, Bhabha Engineering Research Institute (BERI), Bhopal, MP, India

[2]Dean (R&D), SIRT, Bhopal, MP, India

Corresponding Author Email: Prof. (Dr.) Mohit Gangwar (mohitgangwar@gmail.com)

ABSTRACT:

Location analytics is key to get into the current market. It deals with movements data of customer which tells about the customer's lifestyle, their behavior through understanding the visiting pattern. A geo-fence is a virtual boundary for a real-world geographic area. A geo-fence can be generated as per user's need by considering different radius, or a geo-fence can be a predefined set of boundaries. Geo-fencing is use of geo-fence. Geo fencing is a feature that uses the global positioning system (GPS) to define geographical boundaries. It has widely increasing scope. Very few applications are dealing with geo-notifications that are intended to inform the mobile/stationary user proactively about location specific information. The technology behind proactive location-based services is called Geo-fencing.

INTRODUCTION:

Location Intelligence is the methodology for transforming location data into business outcomes. Location data can be anything from addresses and latitude/longitude coordinates, to existing points, lines, and polygons. Spatial data can be created with place names and administrative units such as countries and states. It is the capability to understand complex relationships between location and data in order to drive better business decisions. It enables organizations to derive meaningful insight by harnessing the power of location to solve business problems, deliver location-based services or manage assets. In other words, it can also be defined as a process of deriving meaningful insight from geospatial data relationships to solve a particular problem.It involves layering multiple data sets spatially and/or chronologically, for easy reference on a map, and its applications span industries, categories and organizations.

The term "Location Intelligence" is often used to describe the people, data and technology employed to geographically "map" information. These mapping applications can transform large amounts of data into color-coded visual representations that make it easy to see

trends and generate meaningful intelligence. The creation of location intelligence is directed by domain knowledge, formal frameworks, and a focus on decision support. Location cuts across through everything i.e. devices, platforms, software and apps, and is one of the most important ingredient of understanding context in sync with social data, mobile data, user data, sensor data. Location intelligence is also used to describe the integration of a geographical component into business intelligence processes and tools, often incorporating spatial database and spatial OLAP tools.

AIM OF THE STUDY:
Handling the geo data is really challenging task which is quite linked with the concept of big data Analytics. It is the often-complex process of examining large and varied data sets - or big data - to uncover information including hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make informed business decisions. With today's technology, it's possible to analyze the huge 5 data and get answers from it almost immediately – an effort that's slower and less efficient with more traditional business intelligence solutions. Big data analytics helps organizations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations, higher profits and happier customers. Each phase of big data analytics takes too much time and required efficient skill to understand and get the insight from the data. In analytics, people are mostly bother about getting the insight which can be done only with required relevant data. Hence the idea to automate a process which can reduce the time with huge margin, also which require less man power and automate few most frequent analytics requirements like frequency analysis, wallet share analysis, loyalty analysis etc.

Geo Fencing:
 A geo fence is a virtual fence or a perimeter around a physical location. It is a tool to connect the virtual experience with the offline world's physical location.
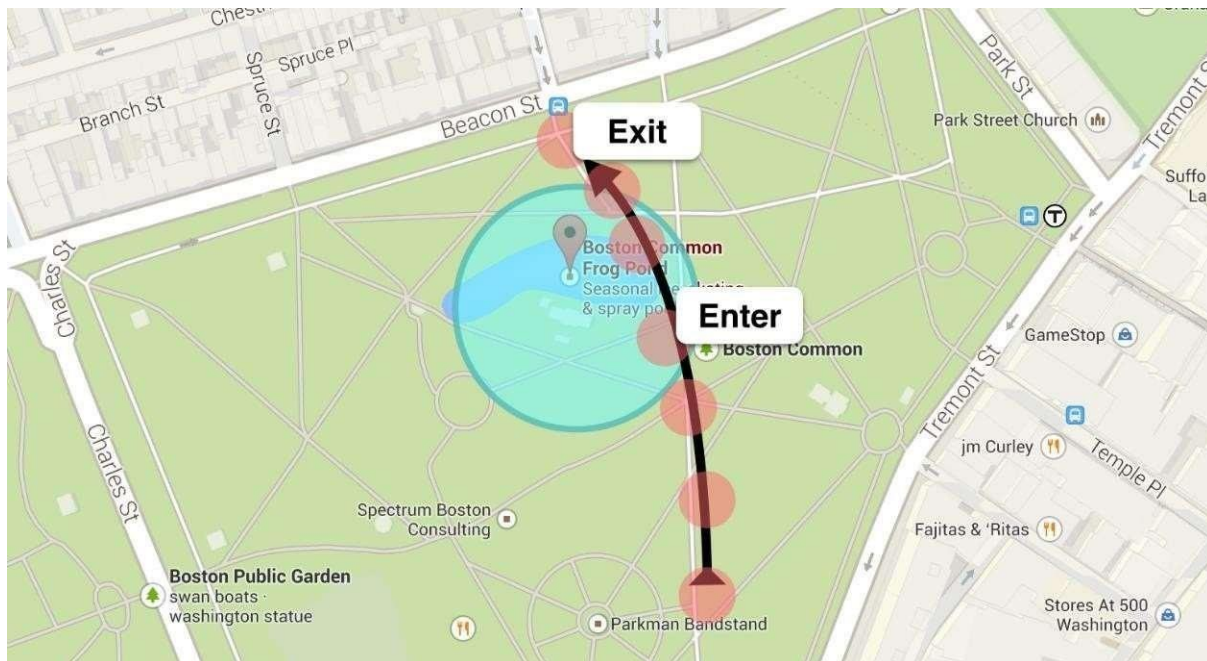
Figure 1.1 Entering and exiting a geo-fence

A geo-fence is a virtual boundary for a real world geographic area. A geo-fence canbe generated as per user's need by considering different radius, or a geo-fence can be a predefined set of boundaries. Geo-fencing is use of geo-fence. Geo fencing is a feature that uses the global positioning system (GPS) to definegeographical boundaries. It has 6 widely increasing scope. Very few applications are dealing with geo-notifications that are intended to inform the mobile/stationary user proactively about location.

RESEARCH METHOD :
Here's the typical flow of procedures for obtaining the user location: 1. Start application. 2. Sometime later, start listening for updates from desired location providers. 3. Maintain a "current best estimate" of location by filtering out new, but less accurate fixes. 4. Stop listening for location updates. 5. Take advantage of the last best location estimate. Below figure demonstrates this model in a timeline that visualizes the period in which an application is listening for location updates and the events that occur during that time. Figure 1.3 A timeline representing an application listens for location updates. [5] This model of a window—during which location updates are received—frames many of the decisions you need to make when adding location-based services to your application. [5] 1.6 Challenges in determining user location Obtaining user location from a mobile device can be complicated. There are several reasons why a location reading (regardless of the source) can contain errors and be inaccurate. Some sources of error in the user location include:

● Multitude of location sources GPS, Cell-ID, and Wi-Fi can each provide a clue to user's location. 9 Determining which to use and trust is a matter of trade-offs in accuracy, speed, and battery-efficiency.

● User movement Because the user location changes, you must account for movement by re-estimating user location every so often.

● Varying accuracy Location estimates coming from each location source are not consistent in their accuracy. A location obtained 10 seconds ago from one source might be more accurate than the newest location from another or same source These problems can make it difficult to obtain a reliable user location reading. This document provides information to help you meet these challenges to obtain a reliable location reading. It also provides ideas that you can use in your application to provide the user with an accurate and responsive geo-location experience. 1.7 Limitation in geo fencing Indoor setup – In case you are interested in det
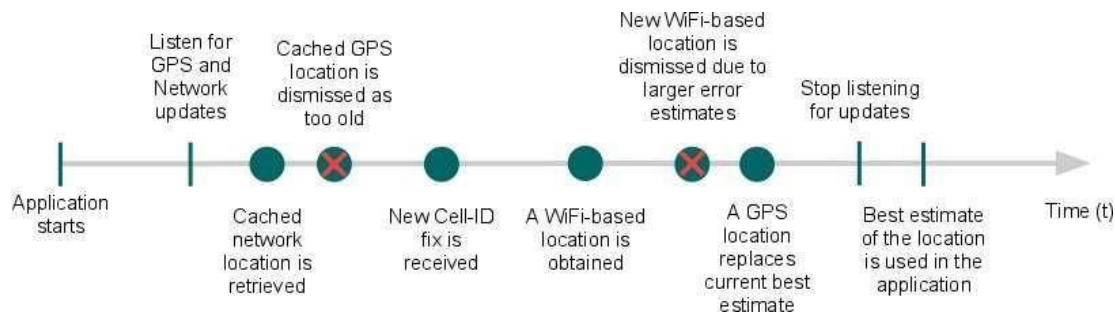

Figure 1.3 A timeline representing an application listens for location updates.

Background functionality –
Geo fencing in the background is crucial to many projects, people generally don't walk with their phone or apps open all the time. Background functionality differs with Android and iOS, but both have options to achieve the desired geo fencing capabilities – you can see our blog post on "Will my geo fencing function in the background" for more details. Size – The size of a geo fence typically raises some questions, how small can I go – or what is the maximum size of the geo fence. There is no limit to how big a geo fence is. You could even geo fence the entire Earth, but then you could just send the same notification to all the app users in bulk, regardless of their position. Geo fencing entire countries, cities or neighborhoods is a perfectly valid use case. 10 The minimum size depends on the technologies utilized, and also a bit on the physical location and the accuracy of GPS/Wi-Fi data there. If you are just using native positioning, a good way to test the minimum size possible for your geo fence by walking around in the area and look at the positioning accuracy in Google or Apple Maps. That is the same native positioning data as your application will be utilizing. In some locations the minimum recommended radius could be 20-50 meters, sometimes you could go over lower. In no circumstance you should go under 5 meters radius with native positioning, as the ability to trigger it comes unstable. When beacons or other indoor positioning setup are added in the picture, you could go as low as 2 meters. This size would ensure there is no unnecessary bouncing in and out of the geo fence. A common use case like geo fencing

different rooms, is quite nicely achievable. Shape – A standard geo fence is circular in shape, but polygon shape geo fences are also possible to create. There are some limitations to what shape is recommended – for example when you want to add a geo fence around a single beacon, circle is the optimal shape. It is also worth noting that the calculating that is required for detecting polygonal geo fences is much heavier, so it will not be as fast or reliable as with circular shapes. Battery consumption – When building your application, you will need to plan for a careful balance between the accuracy, background functionality and battery consumption. Unavoidably, the battery consumption in your app will increase a bit when you activate location services, but with well-designed positioning stack the increase will not be disturbing to your app users.
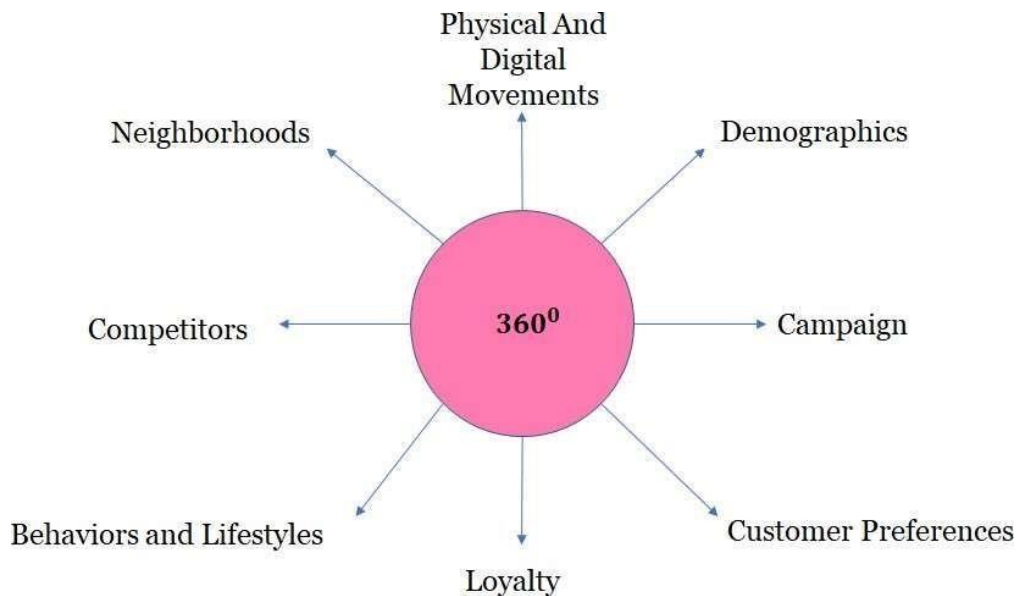

Figure 1.4 Location Analytics - A 360 degree approach

DATA SET AND TECHNIQUES:

Data automation is the process of updating data on open data portal programmatically, rather than manually. Automating the process of data uploading is important for the long-term sustainability of open data program. Any data that is updated manually risks being delayed because it is one more task an individual has to do as part of the rest of their workload. There are three common elements to data automation: Extract, Transform, and Load, or ETL.

● Extract: the process of extracting the data from one or many sources systems

● Transform: the process of transforming the data into the necessary structure, such as a

flat file format like a .csv or .Parquet

● Load: the process of loading the data into the final system.

In fact the term ETL covers the wide range, however this project covers these steps in a programmatically way. In this project, all these three steps of automation are combined in a single Jupiter script.

Data set and outline Before discussing the steps in detail, let's discuss bit about the data set. The data set contains the information of movements of devices and these details are provided by the third party vendor. The set includes device key, location key, latitude, longitude, zip code, state, local date, local time, time zone, brand, movement source key etc. Device key This basically represents the mobile devices Location key This represents a location of a store of a particular brand Latitude and Longitude It represents the exact location of a store Zip Code It identifies the area of the store Local date, time zone It helps to identify the exact time when customer visits at store 15 Brand This is nothing but the name of different bands Source Key It represents from which source data has been collected. e.g. Web, Ads or Wi-Fi etc. Dwell Time It is the time spent by customer in a particular location These device data are added in the data mart on monthly basis hence size of the data is in billions. Due to privacy concern of the company, complete data set details can't be described here, however the above description is quite enough to understand the data set. The outline of the project can be described through the below flow diagram

Data extraction is the act or process of retrieving data out of data sources for further data processing or data storage. In this stage, required data is extracted from geo data mart that will be used for analytics in 3rd phase. For extracting the data in an efficient way, some standard techniques have been used. The complete script has been written in Spark platform and in Spark environment there is a concept of lazy loading. Lazy loading in Spark means that the execution will not start until an action is triggered. In 16 Spark, the picture of lazy evaluation comes when Spark transformations occur. Transformations are lazy in nature meaning when some operation is called in RDD, it does not execute immediately. A sanity test or sanity check is a basic test to quickly evaluate whether a claim or the result of a calculation can possibly be true. It is a simple check to see if the produced material is rational. In this project we have done three checks – Employee check, Panel check and zip code availability check.
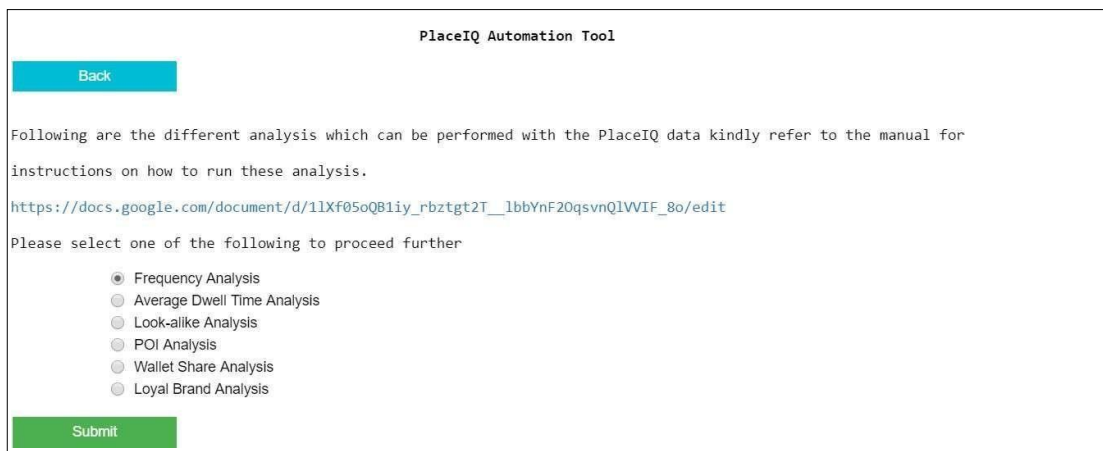
RESULTS AND DISCUSSION:

Analytics is the discovery, interpretation, and communication of meaningful patterns in data; and the process of applying those patterns towards effective decision making. In other

words, analytics can be understood as the connective tissue between data and effective decision making, within an organization. In this chapter, basically 6 different analytics have been discussed that are based on the data after sanity checks.

- Frequency analysis

- Average dwell time analysis

- Look-Alike analysis

- POI analysis

- Wallet share analysis

- Loyal brand analysis

Once you select the third option you will get below screen in your system:

Let's select frequency analysis and click on submit.

4.1 Frequency analysis

Frequency analysis helps to analyze the no. of visits count of a customer's visiting to a brand. This automation tool helps to achieve this analysis at different level –

Level		
Level	• Monthly	This calculates visit frequency for each month in the given time span
Level	• Quarterly	This calculates visit frequency for each quarter in the given time span
	• Store Level	This calculates visit frequency for each stores in the given time span
Store Level	☐ Monthly	This calculates visit frequency for each months and stores in the given time span
Store Level	☐ Quarterly	This calculates visit frequency for each quarter and stores in the given time span

PlaceIQ Automation Tool

Back

You have recently selected:

Selected Checks : Frequency Analysis

Please enter the location of the file generated in step I/II, for more details kindly refer to the documentation:

https://docs.google.com/document/d/1Pyy9wRMD_gaTX1_fdh6I-jSXXXdsV4KMwYdOsPAryjw/edit

Example:- hdfs://10.1.2.138:8020/tmp/USFS/Rapid_PoC/Training/Abhishek/final_filtered_visit_piq.parquet

File: Enter location of the file

Once you select the frequency analysis you will get the below screen to provide the file name on which you need to do this analysis. Input file for this analysis is the output file of sanity check.

CONCLUSION:

The automation work really expedite the process of data extraction. Earlier it takes 2-3 days, just to extract the relevant data. This tool extract the required data in more efficient way. Now it takes almost 30-40 minutes to extract data for five months including 5 states and 5 different brands which is a tremendous improvement. The most important thing is that no manual effort or skill set is required for extraction. User just need to run the script and choose the required option as per their requirement. On the other hand the most frequent analytical component is also added through which any one can get better insight

about the extracted data and take some decision accordingly. In addition to location analytics, the generic data profiling tool kit is also very impressive that helps in exploratory data analysis and cleansing a data set. This helps the user to understand the advance summary of data set and relation between different features through visualization. The advance summary includes - variables, their data type, no. of unique value, no. of missing value, no. of duplicate value, mean, median, mode, standard deviation, minimum value and maximum value for each variables. This notebook will also be helpful in handling the missing value in the data set through different statistical measures like mean, median, mode, zero etc. It also facilitates you to provide your own value for missing value treatment. It also keeps the record of all the changes what user do while performing the missing value treatment. In addition to the features that are already present in the present version there are certain other features which are to be implemented in upcoming versions as mentioned below:

• Linking this notebook to web page for better visuals and interaction.

• Adding visualization and exploratory data analysis features to the notebook , which can help the user take a peek at the data before using it

• Adding other analytical components in the third step. Some of them being, yearon-year analysis, time of day analysis etc.

REFERENCES

[1] Civic media, thinking about the Design with geo–fences, erhardt, April 3rd, 2014, https://civic.mit.edu/2014/04/03/thinking-about-design-with-geo-fences/

[2] Sachin W. Rahate, Dr. M.Z. Shaikh, Geo-fencing Infrastructure: Location Based Service, International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 11 | Nov -2016

[3] Kalle kaitalah, Business development representative in Proximi.io. , https://proximi.io/geofence-complete-guide-geofencing/, January 31st, 2018

[4] https://developer.android.com/guide/topics/location/strategies

[5] https://www.placeiq.com/

[6] Kara Goetz, Data Automation overview, https://support.socrata.com/hc/en-us/articles/212871018-Data-Automation-Overview, March 01, 2019

[7] Yann Moisan, Spark performance tuning from the trenches, https://medium.com/teads-engineering/spark-performance-tuning-from-the-

trenches-7cbde521cf60, May 29, 2018

[8]   Karimi, J., and Walter, Z. The role of dynamic capabilities in responding to digital disruption: A factor-based study of the newspaper industry. *Journal of Management Information Systems*, 32, 1 (2015),

[9]   CPG Matters, "P&G Aims to Enhance Consumer Understanding With New CRM Platform", November 2015

[10]   Javier de la Torre and Santiago Giraldo, Location Intelligence For Dummies®, CARTO Special Edition, John Wiley & Sons, Inc., 2017

[11]   2 Pew Research Center, "The state of privacy in America: What we learned", January 2016

[12]   [12] A. B. M. Musa and J. Eriksson, "Tracking Unmodified Smartphones Using Wi-Fi Monitors," in Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, ser. SenSys '12, 2012, pp. 281–294.

[13]   C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis, and Z. Yan, "Semantic Trajectories Modeling and Analysis," ACM Comput. Surv., vol. 45, no. 4, pp. 42:1–42:32, Aug. 2013. [Online]. Available:

[14]   "Internet Trends D11 Conference," May 2013, http://www.kpcb.com/blog/2013-internet-trends, Accessed: 2015-04-25.

[15]   J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," in Proceedings of the 6th Conference on Symposium on Opearting Systems Design & Implementation - Volume 6, ser. OSDI'04, 2004, pp. 10–10.

[16]   H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 37, no. 6, pp. 1067–1080, Nov 2007.

[17]   M. Naphade, G. Banavar, C. Harrison, J. Paraszczak, and R. Morris, "Smarter cities and their innovation challenges," Computer, vol. 44, pp. 32–39, June 2011.

[18]   J. V. Winters, "Why are smart cities growing? who moves and who stays," Journal of Regional Science, vol. 51, no. 2, pp. 253–270, 2011.

[19]   Thomas, N. O., Singh, S., & Gangwar, M. (2023). Customer retention using loyalty cards program. International Journal of Business Innovation and Research, 30(2), 200-217.

[20] Tembhurne, Y. K., Kiran, T. R., Singh, P., & Gangwar, M. (2024). Solar Still Extraction of Medicinal and Aromatic Plants: Yield, Quality, and Efficiency Analysis. African Journal of Biomedical Research, 27(3S), 2420-2428.

[21] Gangwar, M. (2024). INTELLIGENT COMPUTING APPLICATIONS IN LINGUISTICS.

[22] Gangwar, M. (2024). RESEARCH ASPECTS IN WASTE MANAGEMENT AND THE USE OF INTELLIGENT TECHNOLOGIES.

[23] Gangwar, Mohit (2024). Harnessing Intelligent Computing for Economic Forecasting: Development, Implementation, and Analysis of Advanced Prediction. Rabindra Bharati University: Journal of Economics (2024):61-66.