**Geospatial Capacity Analysis:**

* **The scope of the analysis is to estimate the potential capacity (revenue) that a location can generate for the client. For this purpose we need to assimilate and synthesize multiple datasets from a variety of sources in order to help with the econometric computation of capacity.**

**Geospatial Architectural outline**:-

**S3**:- Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. This means customers of all sizes and industries can use it to store and protect any amount of data for a range of use cases, such as websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics. Amazon S3 provides easy-to-use management features so you can organize your data and configure finely-tuned access controls to meet your specific business, organizational, and compliance requirements. Amazon S3 is designed for 99.999999999% of durability, and stores data for millions of applications for companies all around the world.  
**Advantages of S3**:-  
1. Industry-leading performance, scalability, availability, and durability

2. Wide range of cost-effective storage classes

3. Unmatched security, compliance, and audit capabilities

**EC2** :- Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud computing easier for developers. Amazon EC2’s simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon’s proven computing environment.

**Advantages of EC2:-**

1. Reliable, Scalable, Infrastructure On Demand

2. Vast Breadth and Depth of Compute Services

3. Unmatched security

**Geospatial analysis processes:-**

**Generic Requirements**:-

1. Access to AWS EC2( Ask Chaitra to create AWS credential for any of the new team member)
2. For data upload and download from local system to S3 install an application called ‘S3 Browser’ and ask Chaitra to provide you with IAM access credentials to login into S3 Bukcet.
3. Download QGIS in your local desktop. Install library Geopandas, pandas , OS and rtree in your Qgis python console (For Local use)
4. Install Anaconda in your local system and after installation open Anaconda instance to install basic libraries (e.g. Geopandas) to run some of the Geospatial analysis locally.(For Local use)

* **Telecom Data creation**:- This is an iterative process of data aggregation. As per our understanding of retail business we thought of some groupings for the aggregation of Foot traffic data. This concept supports our idea to get better correlation among independent variables and sales data at TZ/Store revenue level.

Steps for data creation:-

1. Data ingestion: We receive data from Telco in their own S3 Bucket or FTP and we get it downloaded either in our local system or AWS EC2 and then we transfer this data to S3 Storage bucket.
2. Data manipulation:- We utilize data from S3 storage later on by reading it using python and further manipulating it according to the groupings(e.g. group by location, city, gender and age) we have created.
3. Data validation:- We store the aggregated output from python code into excel or csv format. To validate if the data produced is correct we cross verify the numbers produced in different data segmentation.

* **Heatmap creation**:- In this process we created hexagonal shapes based upon the location coordinate shared by Telco. This is done basically to generate the location based heatmaps for foot traffic data

**How to create polygon Heatmaps in QGIS**

Upload the points data as vector layer

1. Open Qgis and select Project-> New(to create a new project)
2. In Qgis goto Layer menu -> Add Layer -> Add Delimited Text Layer.
3. Choose the points data file(file.csv) and click on add button.

Upload python file to create Hexagon

1. Use a shortcut “ctrl+alt+p” to open python console.
2. Open the python file “Hexagon\_code.py” using open file option.
3. Left click on the newly created layer(Layer we have created in 2nd step)
4. Click the run button to run the python script.
5. Check if Hexagon layer(name = polygon) got created.

Generation of Heatmap in the hexagon layer

1. Right click on the newly created hexagon layer and click on “properties” option.
2. In new window for properties click on “Symbology” option.
3. In the top most dropdown field select “Single Symbol”.
4. In the second top most dropdown list select “Simple fill”
5. In the “Stroke style” fields dropdown select “No Pen” Option.(to avoid borders of Hexagon)
6. In the top most dropdown field select “Categorized”
7. In the “Column” field dropdown choose the Column/Indicator basis which you want to implement the heatmap.
8. Select the color of your choice in “Symbol” and “Color map” fields
9. Click on “Classify” button.
10. Click on “Apply” button to implement all these changes.
11. In the end save project to save all changes you have made

**How to create point Heatmaps in QGIS**

Upload the points data as vector layer

1. Open Qgis and select Project-> New(to create a new project)
2. In Qgis goto Layer menu -> Add Layer -> Add Delimited Text Layer.
3. Choose the points data file(file.csv) and click on add button.

Generation of Heatmap in the Point layer

1. Right click on the newly created hexagon layer and click on “properties” option.
2. In new window for properties click on “Symbology” option.
3. In the top most dropdown field select “Single Symbol”.
4. In the second top most dropdown list select “Simple fill”
5. In the “Stroke style” fields dropdown select “No Pen” Option.(to avoid borders of Hexagon)
6. In the top most dropdown field select “Heatmap”
7. In the “Weight points by” field dropdown choose the Column/Indicator basis which you want to implement the heatmap.
8. Adjust the rendering range to get best quality heat map.
9. Adjust ‘color ramp’ and ‘radius’ options according to your requirement.
10. Click on “Apply” button to implement all these changes.
11. In the end save project to save all changes you have made

* **Spatial Joins**:- this is the exercise we ran as part of analysis on capacity estimate and to get the data for base forecasting model to run the regressions. Primarily we aggregated data for foot traffic, door count, Pois, property data at 2 different levels.  
  1. Trade zone level:- We projected all the different datasets over the trade zone layer and spatially joined these datasets to aggregate all the information present with in Trade zone area to create dataset for Trade zone level capacity estimates. All these datasets were very huge so we used python code to infuse multithreading and multiprocessing in order to process this data in fast pace manner.  
  2. Point level:- This is the done basically to create point of interest level data insights for the client based deliverables. This interest points can be Retail stores, Venues or some other location for which client is providing the geo- coordinates. This was done just to get the aggregates for foot traffic. For this particular analysis we have used distance based spatial joins.

Nearest Neighbour joins:- In this method of aggregation we looked onto the nearest data point for foot traffic to the geo-coordinates our client has provided and merged the data from foot traffic file to point based file for Retail stores , door count, venues etc.

KNN(K nearest neighbours):- In this method of aggregation instead of looking at just one nearest data point for foot traffic to the geo-coordinates, we looked at aggregating the data from multiple nearby points based upon distance from the locations which we want to analyse.