Analysis Performed and Presented by:-

Simran Mishra

IDD Part-II

Roll no. 18064025

Under the guidance and supervision of

Dr. Shishir Gaur

Department of Civil Engineering

REPORT

OBJECTIVE:

This study seeks to examine and analyze the spatial and temporal patterns of 2019 novel coronavirus disease (COVID-19) outbreaks and identify the spatiotemporal distribution characteristics and changing trends of cases.

Emerging spatiotemporal hot spot analysis is performed to analyze the spatiotemporal clustering pattern and cold/hot spot trends of COVID-19 cases based on space-time cube.

INTRODUCTION:

Throughout the development of human society, infectious diseases have been a public health problem that cannot be overlooked. These diseases have posed serious threats to human health and socioeconomic development.

Studies found that COVID-19 is caused by the novel coronavirus, SARS-CoV-2.

- It is a novel betacoronavirus belonging to the lineage B (subgenus: sarbecovirus), which also contains SARS-Cov.5
- Genetic analysis indicated that SARS-CoV-2 is genetically very similar to SARS-Cov and the coronaviruses isolated from the bats, but the symptoms are significantly distinct.
- SARS-CoV-2 spreads mainly through respiratory droplets and contact transmission which is similar to the common flu, but potential feces-oral transmission is possible according to some studies.

Despite efforts in clinical research and epidemiological investigation, we still have no answers to the following questions:

1. What is the range of the full spectrum of disease severity?
2. Who is infectious? Can we identify them?
3. How does the vital interest role of asymptomatic infected individuals play in the transmission?

4. How can we identify potential high-risk groups or areas for prevention and control so that we can focus on optimization resources?

The present study will identify the space-time clustering pattern, determine the cold and hot spot trends, and show the results of the visual output for COVID-19 outbreak through the spatiotemporal analysis technique based on STC to enable observation of the identification of high- and low-risk areas intuitively and provide information that can be used as basis for the next step in prevention and control strategy formulation.

METHODS

(A) Data sources

All data in the present study were gathered from the Internet. The COVID-19 data, including confirmed cases, suspected cases, and deaths were obtained from the Indian Center for Disease Control and Prevention website.

(B) Analysis methods

We have performed the analysis on a model data of Restaurant Check-ins in Bay Area, USA. We will then use this analysis to work on the Covid-19 data (at present due to the paucity of a detailed data, we have only a crude data for COVID-19 in Rajasthan and hence till now it has undergone preliminary analysis)

1. **Analyze the data spatially**

Aggregate check-ins to identify spatial clusters of data.

Steps:

- Open the project:
  Firstly we have to investigate the data's attributes.The **User ID** and **Location ID** fields contain unique IDs for users and locations. We don't have access to a key for these IDs, so these fields aren't useful to determine popularity. The **Check-in Latitude** and **Check-in Longitude** fields provide the data's spatial information, while the **Check-in Time** field provides its temporal information.

- Change the coordinate system
  When analyzing the spatial relationships between features, it's important to ensure that you're using a coordinate system that is appropriate for the data. we’'ll project the data to a projected coordinate system that is focused around the required area. This coordinate system minimizes distortion near the required area, at the cost of increasing distortion in other areas. As we're not focused on areas outside it, this coordinate system is appropriate for our map and data.

- Aggregate check-ins
  To gain more meaningful insight, we'll count the number of check-ins in each area. we'll create a grid of hexagon bins that covers the required Area and use this grid to aggregate check-ins. Then, we'll symbolize the result layer to determine which areas have the most check-ins.
  Tools used- Generate Tessellation tool, Spatial Join tool

- Quantify the significance of the aggregations
  We'll use the Global Moran's I statistic to test whether the patterns in our results are clustered, dispersed, or random.
  Global Moran's I quantifies the spatial patterns of an attribute. Because our original check-in data has no attributes that we can use to determine the density of check-ins, it was necessary to aggregate the check-ins before

running the statistics. The hexagon bins have the **Join_Count** field, which Global Moran's I can quantify.

Calculations

The Moran's I statistic for spatial autocorrelation is given as:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{1}$$

where z_i is the deviation of an attribute for feature i from its mean $(x_i - \bar{X})$, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \tag{2}$$

The z_I-score for the statistic is computed as:

$$z_I = \frac{I - \mathrm{E}[I]}{\sqrt{\mathrm{V}[I]}} \tag{3}$$

where:

$$\mathrm{E}[I] = -1/(n-1) \tag{4}$$

$$\mathrm{V}[I] = \mathrm{E}[I^2] - \mathrm{E}[I]^2 \tag{5}$$

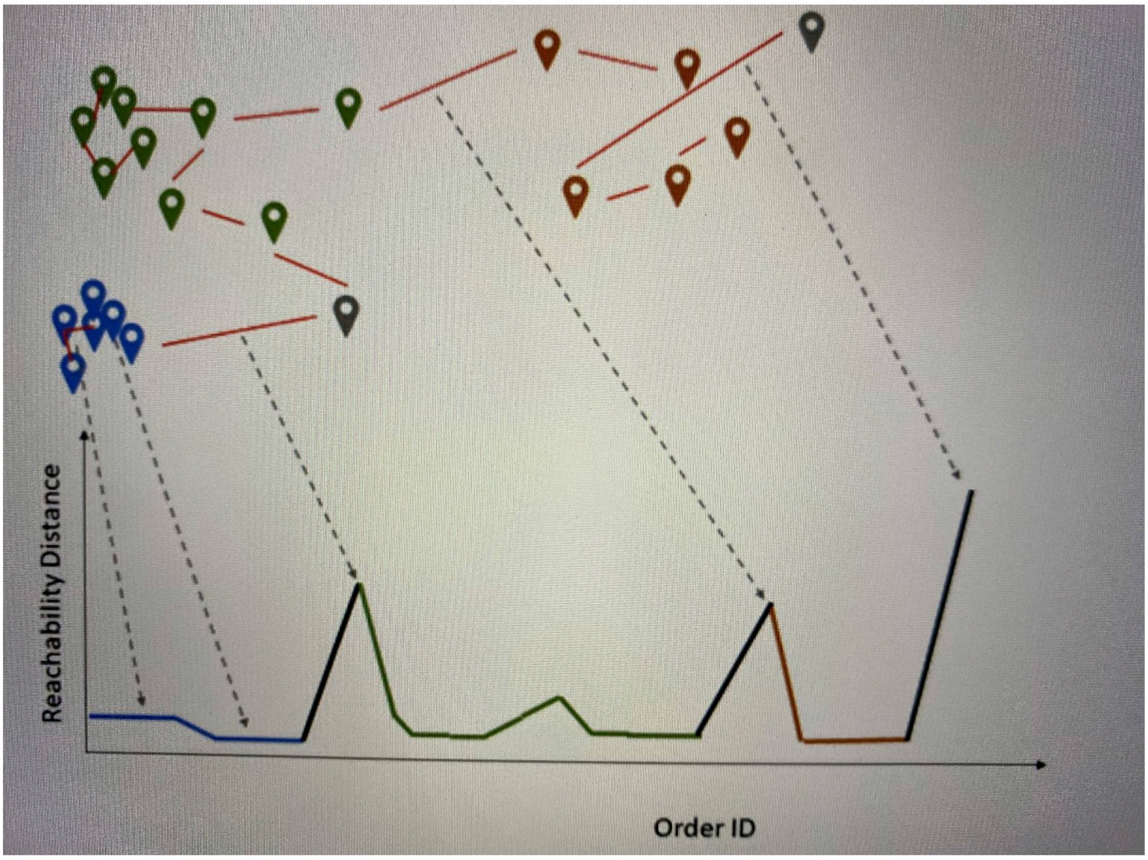
The report includes the Moran's Index, the z-score, and the p-value. For determining statistical significance, the z-score is the most important of these values.

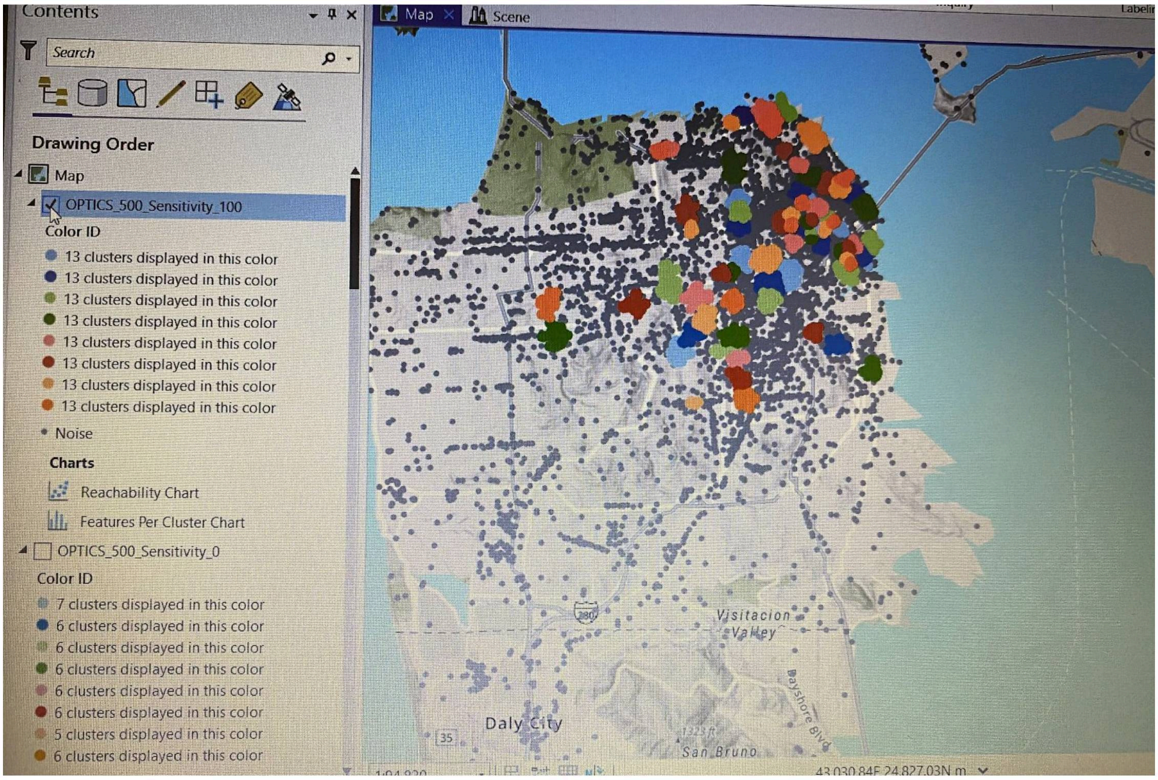
The z-score indicates the number of standard deviations a value is from the average value. Positive z-scores are values above the average, while negative z-scores are values below the average. In this case, the value being measured is the amount of spatial autocorrelation that exists between features in your dataset.

Tools used- Spatial Autocorrelation tool.

- Detect spatial clusters
  By aggregating the data and determining its statistical significance, we know with confidence that check-ins are not randomly distributed, but clustered. Next, we'll perform spatial cluster analysis to detect areas of high concentration of check-ins.

  We've analyzed our data spatially.
  Through aggregation and spatial clustering, we've determined locations where there are particularly high densities of check-ins and learned some of the ways to adjust our analysis results depending on your specific objectives.





2. **Analyze the data temporally**

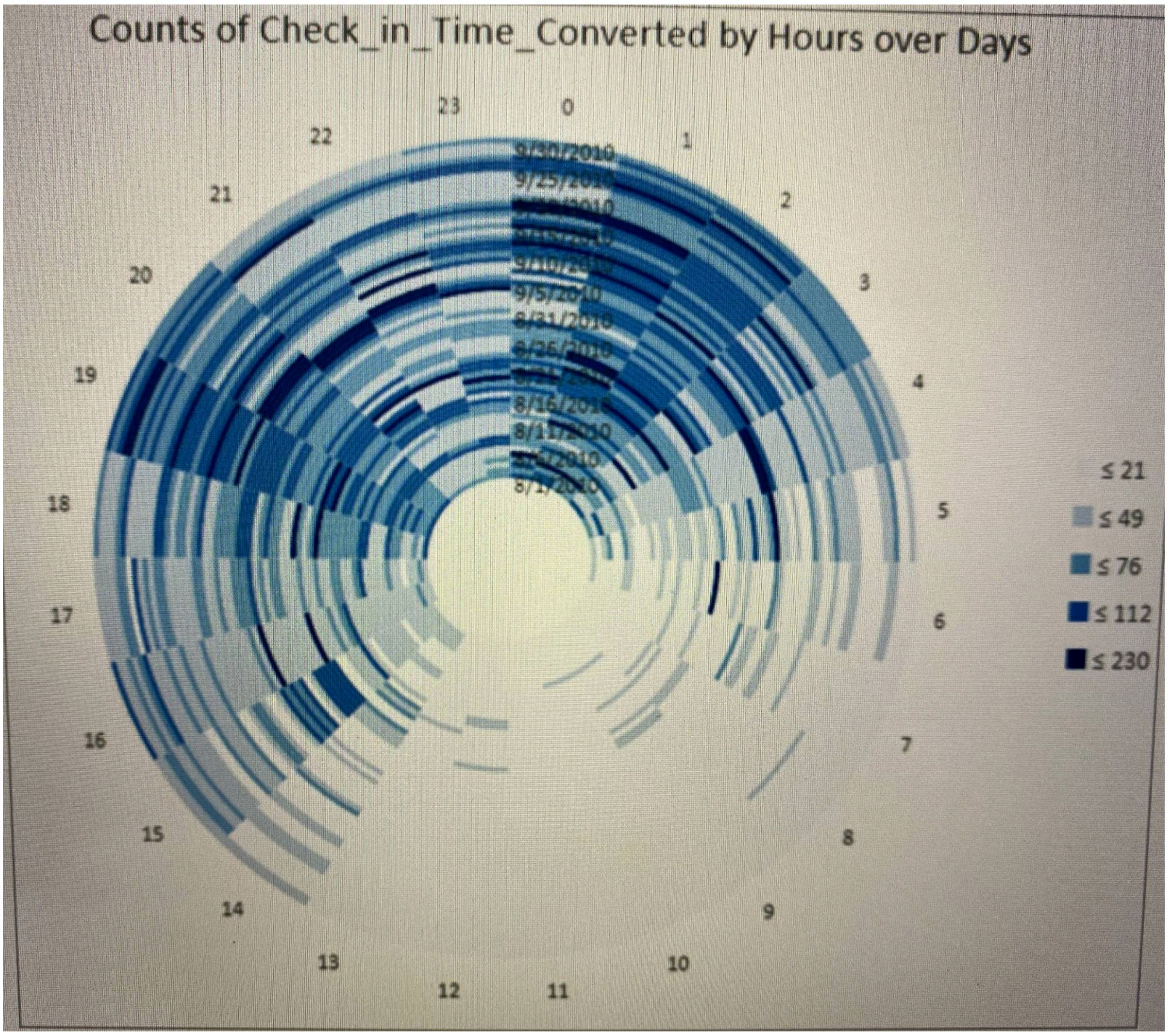
Our data has both a spatial and temporal component. Analyzing spatial trends is useful, but it doesn't tell the entire story. So here we go for temporal analysis.

Steps:

- Convert the time field
  The Check-in Time field contains the date and time when a check-in was created. However, the field contains a concatenated string of text that ArcGIS Pro does not automatically recognize as a timestamp. To use this field for temporal analysis, we'll convert it to a recognized data field format.

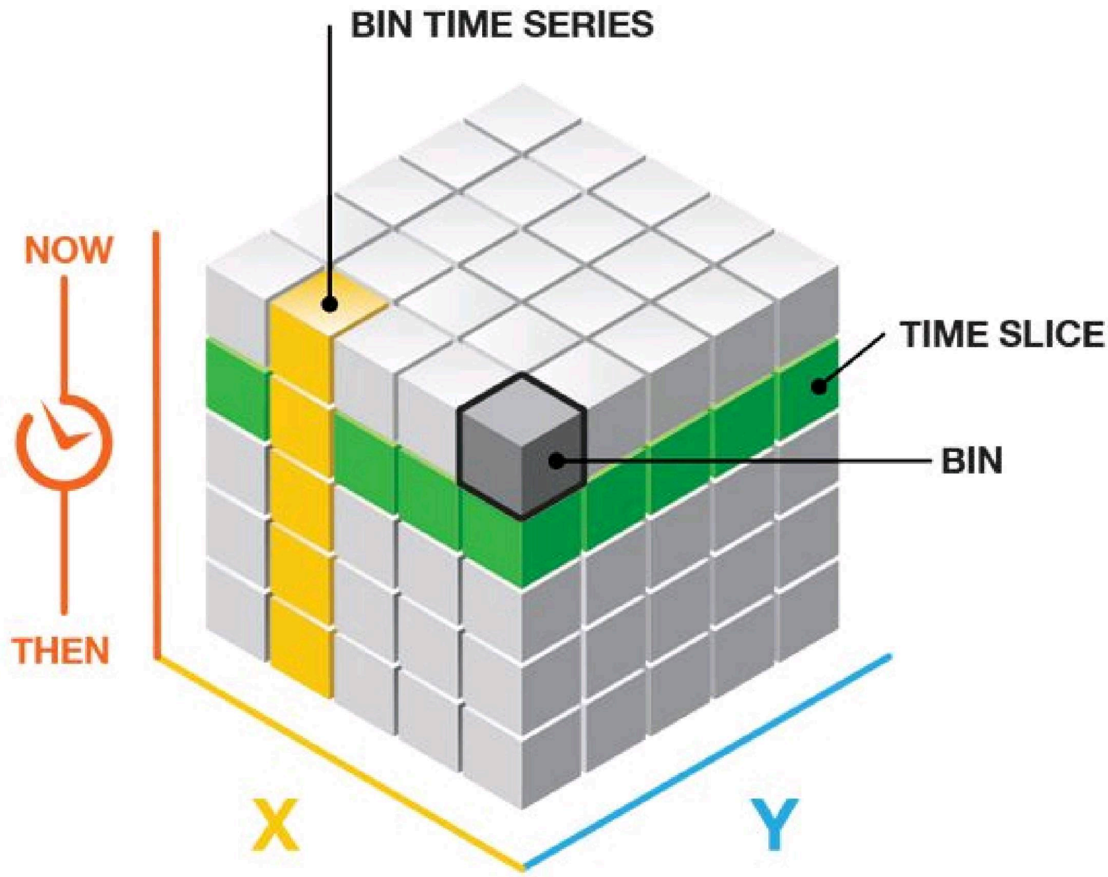
- Chart the temporal data
  Our feature class contains time data that ArcGIS Pro can process and analyze. Next, we'll create a data clock. Data clocks are a type of chart that

summarize temporal data. we'll use this chart to find patterns in the times people checked in.



- <u>Analyze trends with a space time cube</u>
  To analyze the spatial and temporal elements of our data together, we'll need to create a spatiotemporal data structure (a data structure that accounts for both space and time). This data structure will summarize the check-in points by a fixed area and a fixed increment of time.
  we'll use the **Create Space Time Cube** tool to define a spatiotemporal data structure for your data. The resulting dataset can be thought of as a cube

because it has three dimensions: two for area (x and y) and a third for time (t).



- Detect temporal clusters
  Next, we'll detect temporal clusters of check-ins in your space time cube. Temporal clustering is similar to spatial clustering in that it identifies locations where features are densely grouped. The only difference is that temporal clustering groups clusters by temporal proximity instead of spatial proximity.

3. **Complete your analysis**

Upto now we've analyzed our data spatially and temporally. Depending on which statistical method we choose to detect clusters in your data, our results may change significantly. Next, we'll combine our results and reach a decision about hot/cold spots in the study region.

Steps:

- Detect spatial and temporal hotspots

  Our final analysis will examine the data spatially and temporally at the same time. Using Emerging Hot Spot Analysis (EHSA), we'll classify patterns in our space time cube.

  Unlike time-series clustering, EHSA determines whether a space time cube bin's neighbors contain a number of check-ins that are significantly higher than (hot spot) or lower than (cold spot) the global average. Once every location in the space time cube has been designated a hot spot, cold spot, or neither, EHSA examines variations in each location's z-score over time to determine whether the location is a consecutive, intensifying, diminishing, or sporadic hot or cold spot.

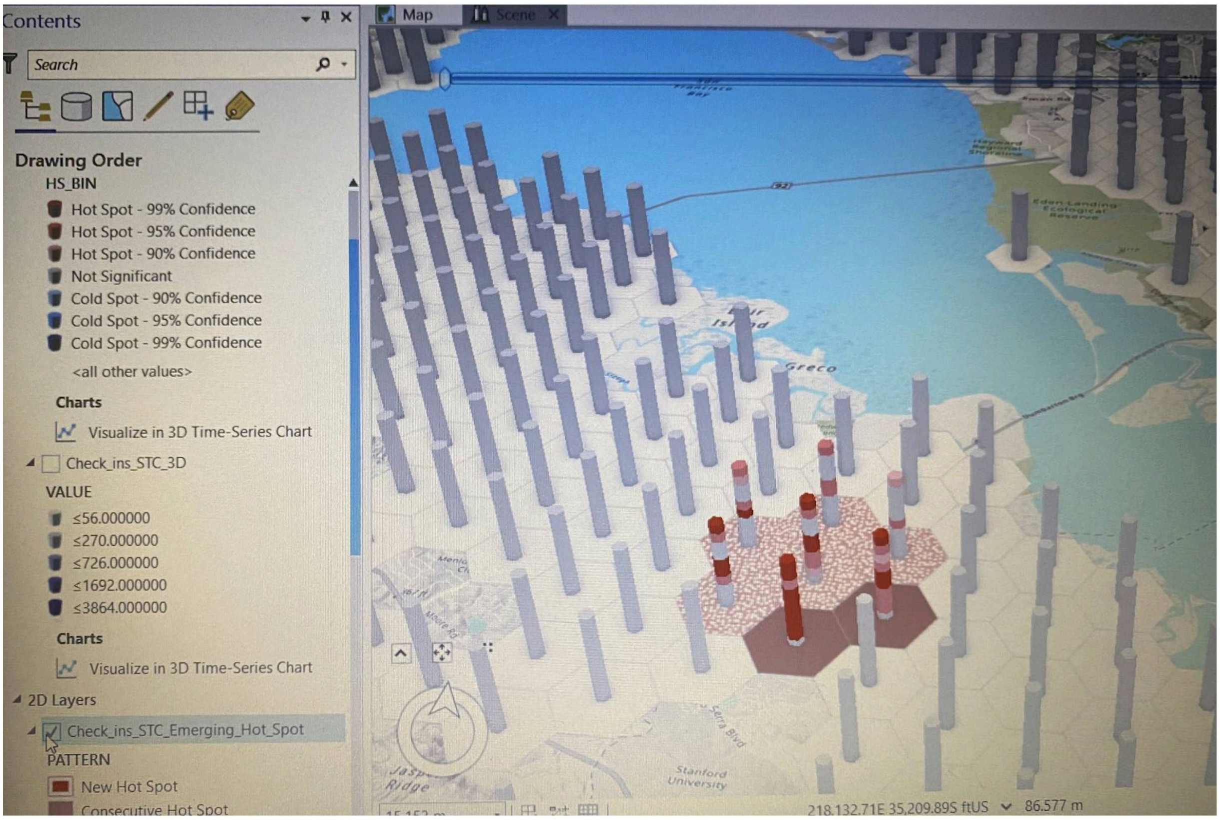
  The final result accounts for both spatial and temporal variations in the data.

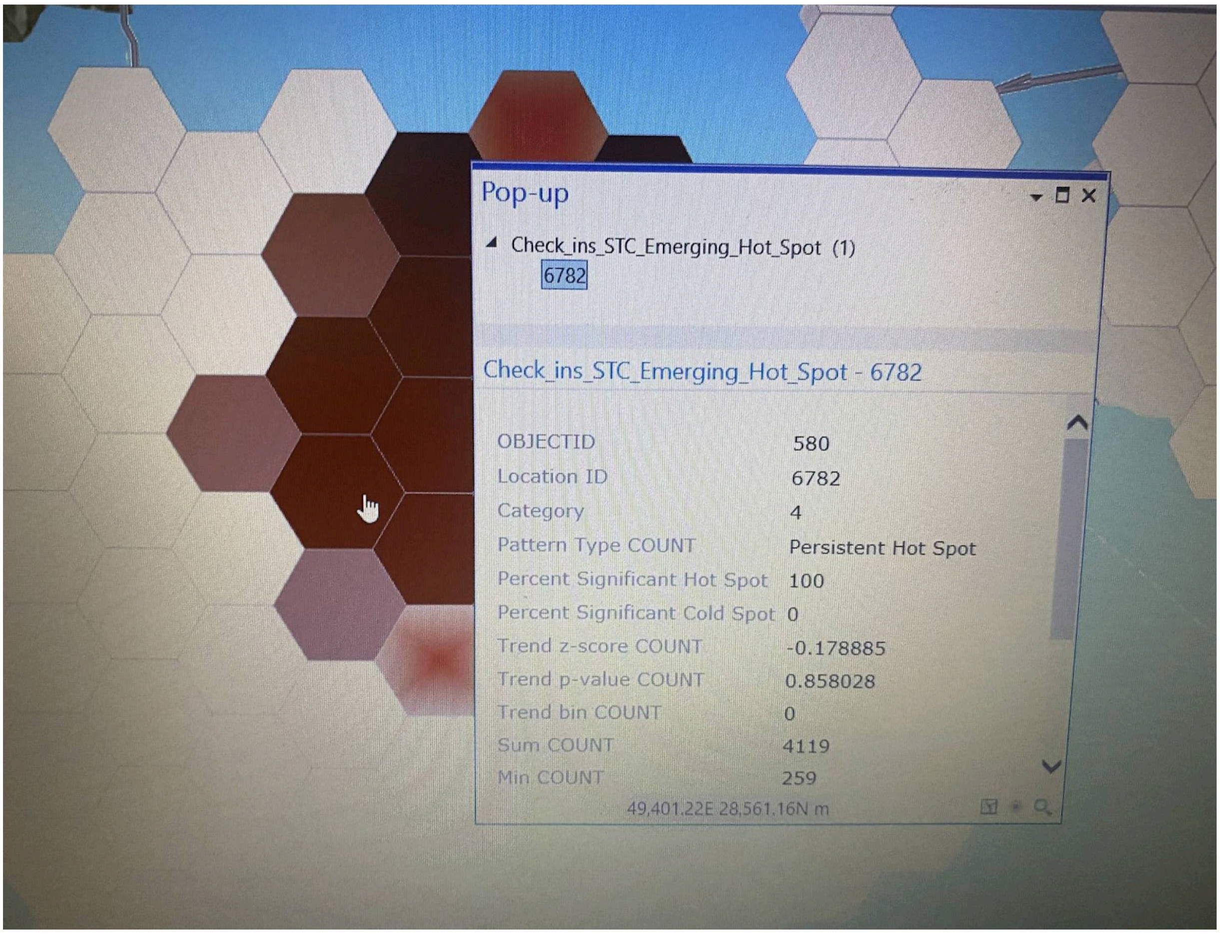
- Decide where to open your business

  Next, you'll determine the best location to open your new business. To do so, you'll overlay your spatial clusters, temporal clusters, and emerging hot spots.

The criteria for how you combine these layers will depend on what you consider the ideal conditions for your business.

First, you'll select areas with dense spatial clusters of check-ins. These areas indicate high foot traffic, which is good for a new business. You performed spatial cluster analysis using three different methods: DBSCAN, HDBSCAN, and OPTICS. Of the three, HDBSCAN was the most appropriate to your study area, as it accounted for differences in population between the Bay Area's urban, suburban, and rural locations.


Contents
Search
Drawing Order
HS_BIN
Hot Spot - 99% Confidence
Hot Spot - 95% Confidence
Hot Spot - 90% Confidence
Not Significant
Cold Spot - 90% Confidence
Cold Spot - 95% Confidence
Cold Spot - 99% Confidence
<all other values>
Charts
Visualize in 3D Time-Series Chart
Check_ins_STC_3D
VALUE
≤56.000000
≤270.000000
≤726.000000
≤1692.000000
≤3864.000000
Charts
Visualize in 3D Time-Series Chart
2D Layers
Check_ins_STC_Emerging_Hot_Spot
PATTERN
New Hot Spot
Consecutive Hot Spot
Map
Scene
Greco
Stanford University
218,132.71E 35,209.89S ftUS
86.577 m



(C) Tools Used

1. **In spatial Analysis:**

- Project tool: To open the project in ArcGIS software.
- Generate Tessellation tool:

  This tool creates a grid of regular polygon features, such as hexagons, squares, or triangles, to cover a specified extent.
- Select Layer By Location tool :

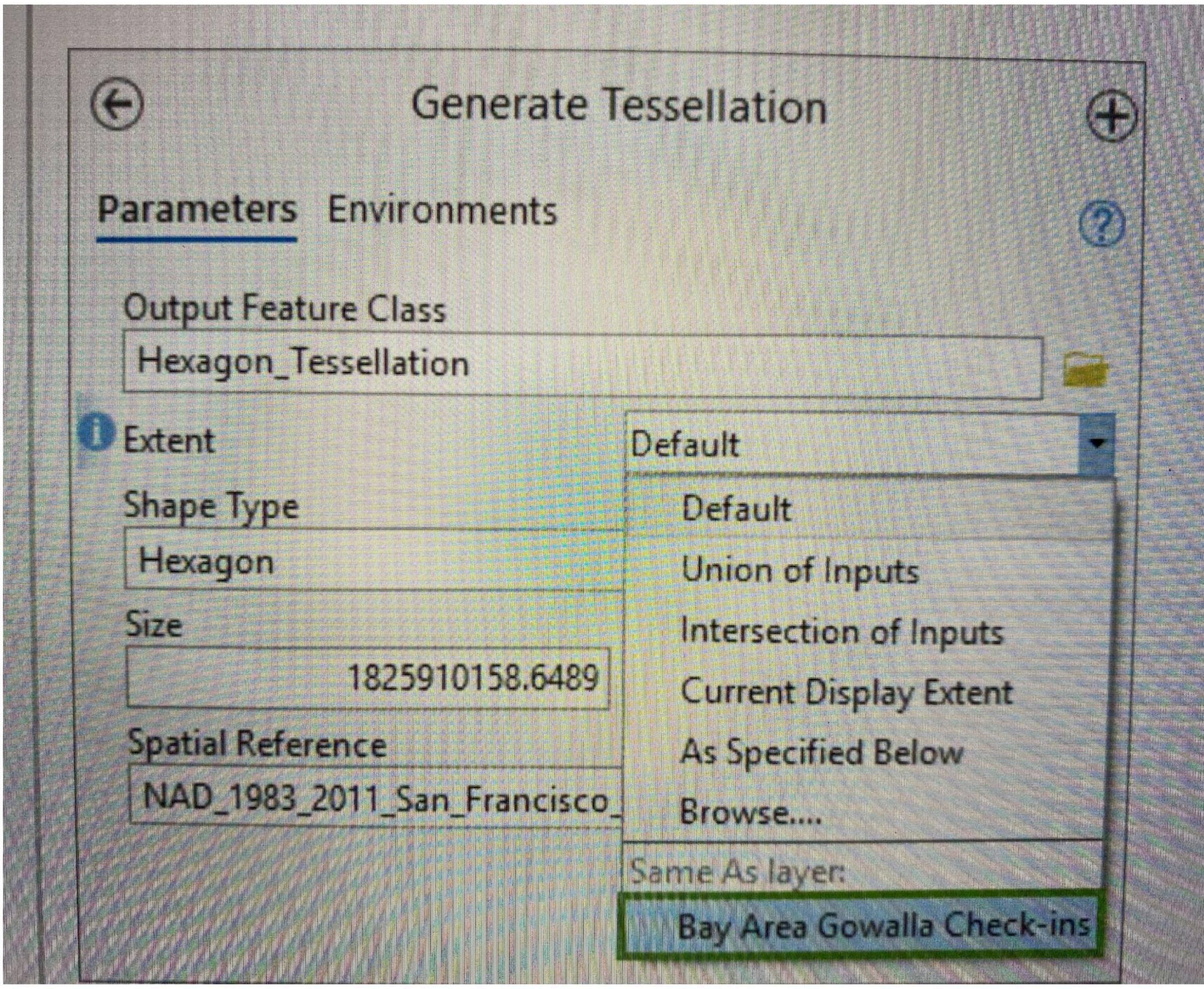
  It opens in the **Geoprocessing** pane.

- Spatial join tool:

  The tool runs and a new layer, containing only the selected hexagon bins, is added to the map. The check-in counts for each bin are contained in an attribute field for the layer. To visualize the counts on the map, we'll change the layer symbology.
- Spatial Autocorrelation (Global Moran's I) tool:

  The tool runs, but no layer is added to the map. Instead, a report file was created. You can find the path to this report file by viewing information about the tool.
- Density-based Clustering tool:
  - ➔ This tool provides three methods for spatial clustering, with each method requiring a different definition of what is considered dense and not dense. We'll run the tool three times, one for each method, and weigh the advantages and disadvantages of each.
  - ➔ First, we'll use the defined distance method, also known as **DBSCAN, which is the simplest density-based clustering method**. In this method, density is defined by having a specified number of points within a specified distance. At every point, it checks whether the point satisfies the minimum number of features within a set search distance. If a point satisfies this criterion, it is marked as a clustered point. To run the tool, you must define the minimum number of features. You can also define the search distance, but if you do not set a search distance, the tool uses an optimized value.
  - ➔ The tool runs and the result layer is added to the map.

- ➔ Compared to the DBSCAN method, the HDBSCAN method detects more clusters. Clusters appear all over the Bay Area, including in rural areas, and some of these clusters are large enough to cover the entire city of rajasthan.,
- ➔ Next, we'll use the third spatial clustering method, multi-scale (also called OPTICS).
- ➔ The OPTICS method records the distance between the first feature in a dataset (Order ID 0) and its nearest neighbor. This distance is called the reachability distance. Then, the method records the reachability distance between the nearest neighbor and its nearest neighbor. This process repeats continually until the entire dataset has been covered. No nearest neighbor is repeated; if one feature's nearest neighbor was also the nearest neighbor of a previous feature, the next nearest neighbor is used instead.



2. **<u>In temporal Analysis:</u>**

- <u>Convert Time Field tool:</u>

  This tool converts time and date values from a text string to a date field.

- <u>Copy Features tool:</u>

  The copied feature class is added to the map.

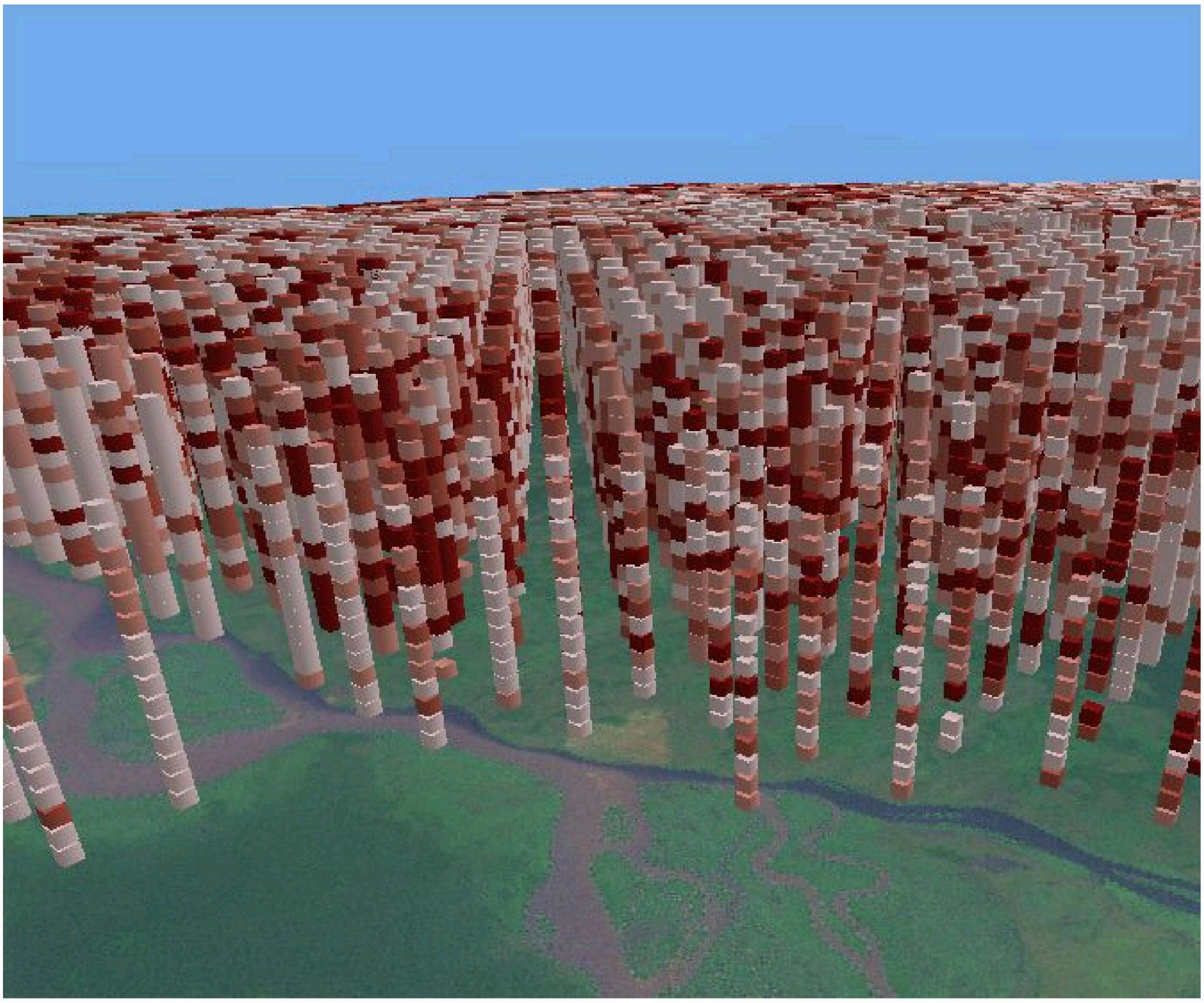
- <u>Create Space Time Cube tool :</u>

  The tool that you choose depends on your data. Your check-in data comes from a variety of point locations across space, so you want to aggregate points. If your data instead relied on stations or other locations with fixed geographies (such as traffic cameras or toll booths), you would create a space time cube from defined locations. If your data came from a multidimensional raster layer, you would choose the appropriate tool.

- Visualize Space Time Cube in 2D tool:
  This tool creates a 2D layer based on an .nc file.
- Visualize Space Time Cube in 3D tool:
  One interesting way to inspect the space time cube data is with the Visualize Space Time Cube in 3D tool. This tool, located in the Utility toolbox within the Space Time Pattern Mining toolbox, creates a three-dimensional representation of the bins that can be viewed in ArcGlobe or ArcGIS Pro.

  In this visualization, each hexagon bin has a height composed of segments, with each segment corresponding to a different week. The color of each segment indicates the number of check-ins in that area during that week..

  Unlike the 2D visualization, each segment is symbolized by total count of check-ins, not by increasing or decreasing trends.



- <u>Time Series Clustering tool:</u>

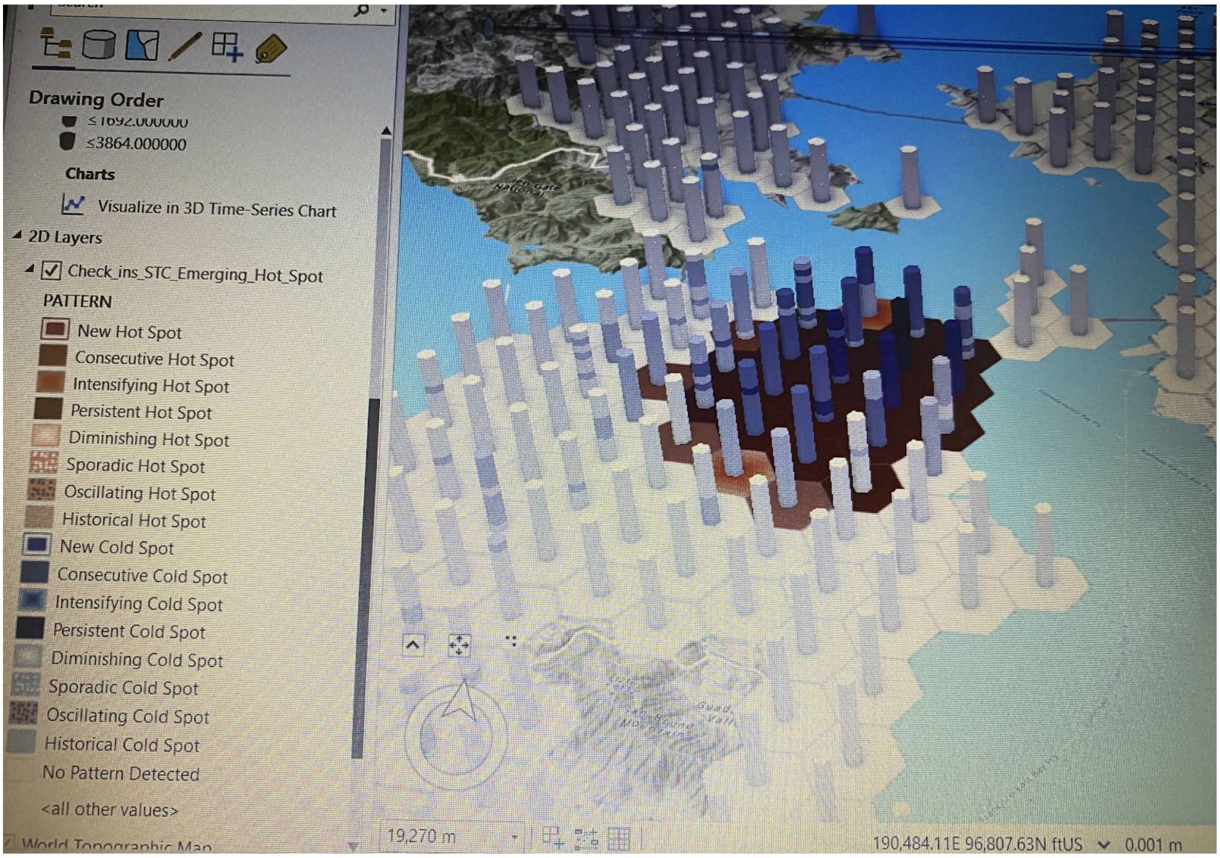
  we can also cluster the data by one of three characteristics of interest. We can also set the number of clusters that the tool creates. If left unchanged, the tool will use an optimal number based on the data. we'll create three clusters, corresponding to groups of high, medium, and low number of cases/deaths of covid-19.

3. **<u>In complete Analysis:</u>**

- <u>Emerging Hot Spot Analysis tool:</u>

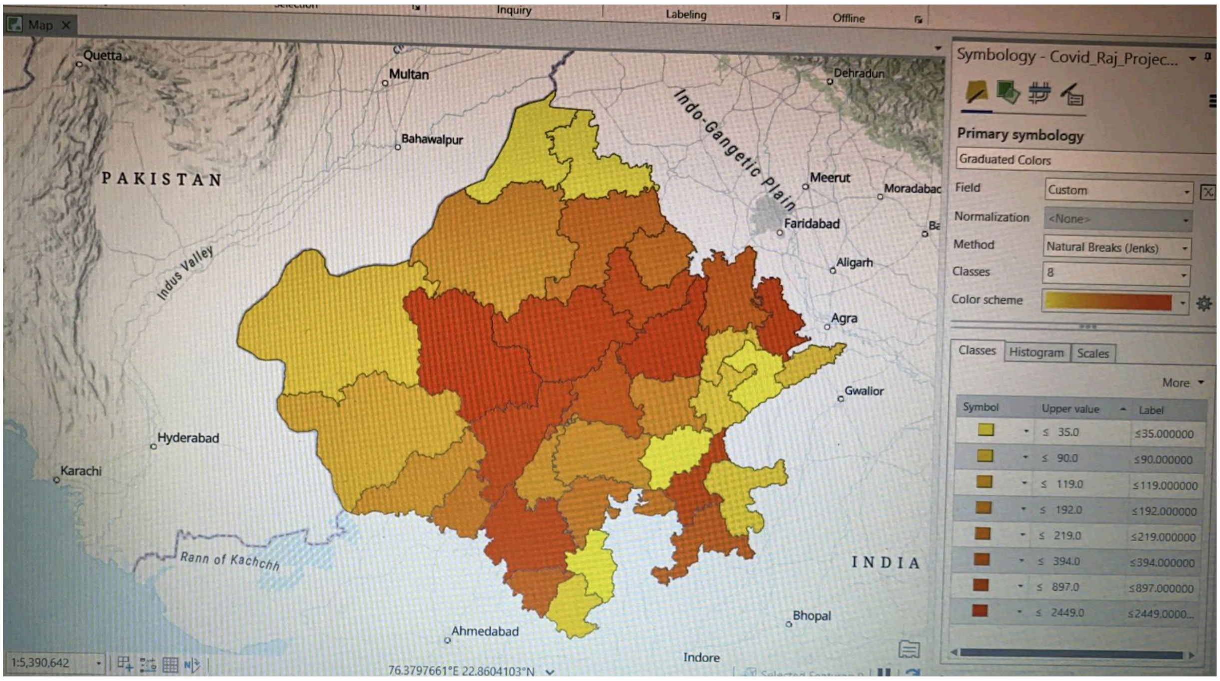
  For each location, EHSA will examine every neighboring location within a mile to perform its analysis. We previously created a space time cube with a hexagon grid, which is ideal for neighborhood analysis because each hexagon is equidistant.
- <u>Select Layer By Attribute tool:</u>

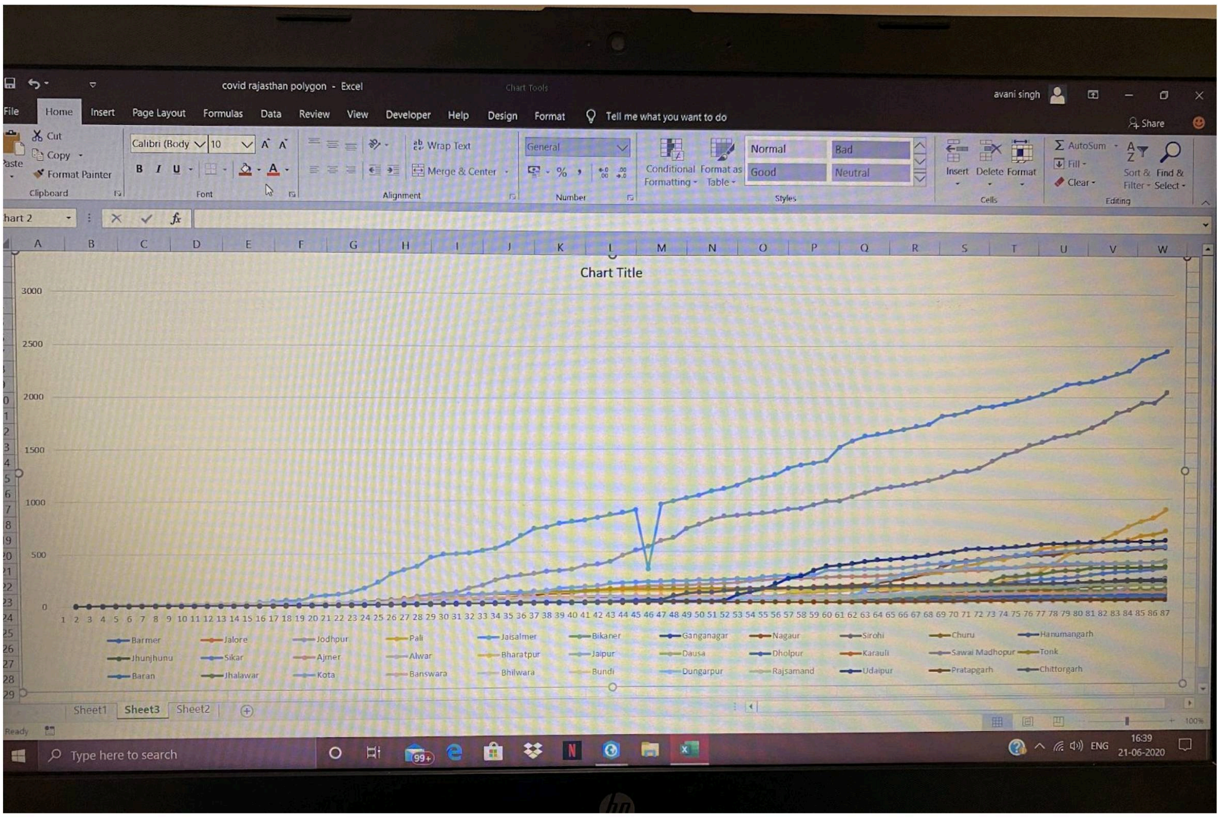
  The hot spots are selected. Next, we'll select weekly time clusters that see an increase in the number of cases/deaths at the particular week. Depending on the type of region either hot/cold spot, areas with more number of cases/deaths will be idle.



<u>Result</u>

In the actual covid-19 data for Rajasthan. Till now we have just reached to a conclusion to which city shows which type of growth trend, with Jaipur and Jodhpur amongst the fastest growing hotspots for Covid-19 as indicated by the fast growing line. The map shows deep red for districts with the highest numbers of total cases till the latest known date.


Symbology - Covid_Raj_Projec...
Primary symbology
Graduated Colors
Field
Custom
Normalization
<None>
Method
Natural Breaks (Jenks)
Classes
8
Color scheme
Classes
Histogram
Scales
Symbol
Upper value
Label
≤ 35.0
≤35.000000
≤ 90.0
≤90.000000
≤ 119.0
≤119.000000
≤ 192.0
≤192.000000
≤ 219.0
≤219.000000
≤ 394.0
≤394.000000
≤ 897.0
≤897.000000
≤ 2449.0
≤2449.0000...
PAKISTAN
INDIA
Quetta
Multan
Bahawalpur
Hyderabad
Karachi
Indus Valley
Rann of Kachchh
Ahmedabad
Indore
Bhopal
Gwalior
Agra
Aligarh
Faridabad
Meerut
Dehradun
Indo-Gangetic Plain
1:5,390,642
76.3797661°E 22.8604103°N

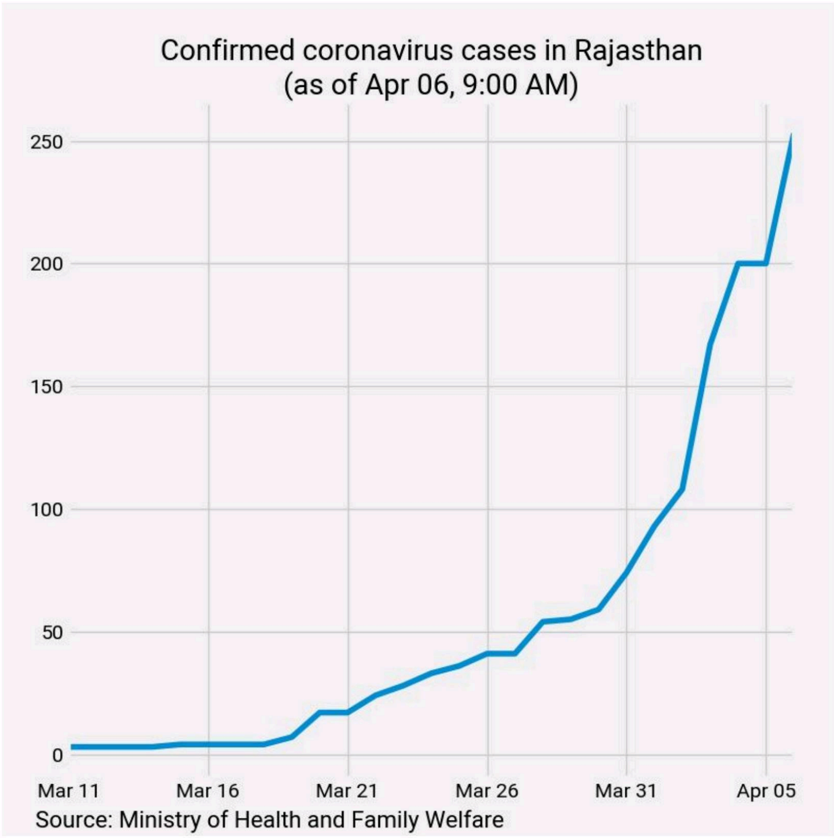

covid rajasthan polygon - Excel
Chart Title
3000
2500
2000
1500
1000
500
0
Barmer
Jalore
Jodhpur
Pali
Jaisalmer
Bikaner
Ganganagar
Nagaur
Sirohi
Churu
Hanumangarh
Jhunjhunu
Sikar
Ajmer
Alwar
Bharatpur
Jaipur
Dausa
Dholpur
Karauli
Sawai Madhopur
Tonk
Baran
Jhalawar
Kota
Banswara
Bhilwara
Bundi
Dungarpur
Rajsamand
Udaipur
Pratapgarh
Chittorgarh
Sheet1
Sheet3
Sheet2

Discussion

- In this study, we analyze and excavate the spatiotemporal pattern of COVID-19 outbreak and determine the high- and low-risk areas of the epidemic by identifying the spatiotemporal clustering pattern and cold or hot spot trends of cases based on STC and spatiotemporal analysis technology. The main findings are as follows.

  (1) The outbreak had spread rapidly throughout the state within a short time and the total incidence rate has decreased.
  (2) The spatiotemporal distribution of cases is uneven.
  (3) The spatiotemporal analysis technology based on the STC can analyze comprehensively the spatiotemporal pattern of the epidemiological data and produce a visual output of the consequences, which can reflect intuitively

the distribution and trend of data in space-time.



References:-

- https://learn.arcgis.com/en/projects/identify-popular-places-with-spatiotemporal-data-science/#analyze-the-data-temporally
- https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm
- https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.25834
- https://www.azavea.com/blog/2017/08/15/emerging-hot-spot-spatial-statistics/