Blue Sky Analytics Assignment

Blue Sky Analytics

This document contains a summary report answering the below questions from the assignment given:

1. Summary of your solution.
2. Key outcomes or takeaways from your solutions.
3. How did your solution involve geospatial data?
4. Does your solution require improvements in terms of data or methodology?

NOTE: Since this is a summary report, no visualisations were included.

Visualisations and more elaborations were made in the “Blue Sky Assignment” slide

Blue Sky Analytics Assignment

Blue Sky Analytics

1. Summary of your solution

My solution:

- Employed India as the regional case study by analysing the most critically hit developing country during the COVID-19 pandemic
- Combined satellite data and ground observations to perform exploratory analysis on the air and water quality of India during the pandemic and correlated between changes in ground activities and their environmental impact
- Involved machine learning to accurately predict the air quality over the course of the pandemic
- Used multiple time-series data and geographic regions to verify my analysis.

Blue Sky Analytics Assignment

Blue Sky Analytics

2. Key outcomes or takeaways from your solutions.

- India was the most critically hit developing country during the COVID-19 pandemic
- Mobility, retail and recreation activities dropped drastically during the pandemic and resumed once more after the lockdown measures were relaxed
- As a result, air quality, which was previously poor, became better during the lockdown, signified by a reduction in the monthly average pollutant levels during lockdown.
- The decreased trend in pollutant levels is correlated to the mobility pattern observed during the lockdown

Blue Sky Analytics Assignment

Blue Sky Analytics

2. Key outcomes or takeaways from your solutions.

- More than half of the signal was due to the lockdown, with only the highest concentrations of NO2 (due to industrial activity) remaining near the **orissa region**, where most of the refinery and industrial capability is located as it would be expected that there would be some industrial activity to support essential services although the activities were far less.
- The air quality degraded once more after the pandemic, with pollutant concentration levels even higher than the levels before lockdown, caused by a great rate of increase in mobility, retail and recreation activities.

Blue Sky Analytics Assignment

Blue Sky Analytics

2. Key outcomes or takeaways from your solutions.

- There was a clear trend that pollution level in India falls in the month of July and August. This might be majorly because monsoon season sets in during these months
- The pollution level then start rising and reach highest levels in winter months.
- I compared the levels for 5 different pollutants this time (NO2, CO, SO2, particulate matter and BTX (Benzene, Toluene and Xylene)).
- The median values of 2020 are generally less as compared to other years giving us a sense that there might be a reduction on pollution during the year of the pandemic, which objectively agrees with the previous analysis.

Blue Sky Analytics Assignment

Blue Sky Analytics

2. Key outcomes or takeaways from your solutions.

- Comparing the Air Quality Index (AQI) of major Indian cities before and after lockdown showed that the air quality improved after lockdown, signified by a lower AQI value
- I then developed a machine learning model to predict the AQI of Delhi from the end of July, 2020 to June 2021.

  The model was able to correctly predict the trend with minimal effort. This would be helpful for evaluating the spread of COVID-19 in the future, which in turn, could be used for evaluating the times and seasons with good or bad air quality and the respective environmental changes.

Blue Sky Analytics Assignment

Blue Sky Analytics

2. Key outcomes or takeaways from your solutions.

- Using remote sensing data, there were improvements in ambient water quality in terms of decreased chlorophyll-a levels for a section of the Tapi River in the Gujrat region of India.
- The closure of anthropogenic and industrial activities, thus resulted in the decline of pollutants concentration in the Tapi River.
- The upsurge in Chlorophyll-a concentration observed may be supported by the fact that the natural river sedimentation can also alter the those levels in water bodies. The flow of sediments is governed by many factors such as precipitation, water and soil conservation projects, reservoirs, etc.

Blue Sky Analytics Assignment

Blue Sky Analytics

2. Key outcomes or takeaways from your solutions.

- An average decrease of chlorophyll-a (NDCI) concentrations during lockdown compared to the pre-lockdown period despite the domestic wastewater continuing to drain into the Tapi River showed that the major deteriorating factor of the river is the industrial sewage discharge and other outdoor activities, which severely declined during the lockdown measures, hence severe pollution stopped

Blue Sky Analytics Assignment

Blue Sky Analytics

3. How did your solution involve geospatial data?

I employed geospatial data for every stage of analysis;

- I applied geospatial analysis using the world map to visualise countries with the biggest hits during the COVID-19 pandemic, thus identifying India.
- I used geospatial geojson data of South Asian countries to analyse and present a heat map of the pollutant levels of NO2 and CO around every region in India
- I set up openstreetmap and stamen toner map views to inspect the Air Quality Index of different cities in India in real-time.
- I used geospatial data to analyse the water quality (NDCI levels) of Tapi river in India

Blue Sky Analytics Assignment

3. Does your solution require improvements in terms of data or methodology?

- Since this was an assignment, I did not expose my ML model to increased training time. Thus, a minor change would be increasing the training time, while possibly adjusting the hyperparameters.
- I handled some missing data from the Air Quality Index data I obtained. While this does not affect the results of analysis, it would be nice if the data were complete.

Blue Sky Analytics Assignment

All questions have been successfully answered.

Feedback or comments would be welcomed, thank you!