School of Computing

AY20/21 Semester 1

CS4225 Big Data Systems for Data Science

Climate, Traffic and COVID-19: Project Report

Project Team 7

GitHub Repository: https://github.com/pratyayj/cs4225-t7-project

Tableau Visualisations: Climate Dashboard and Traffic Dashboard

Table of Contents

1. Project Introduction

In 2020, we have come to see everything through the lens of COVID-19. The pandemic has irrevocably altered the way we live our daily lives. When things get better, it is crucial that we build back better. In the words of the United Nations Secretary-General António Guterres, "Everything we do [...] must be with a strong focus on building more equal, inclusive and sustainable economies and societies."[1]

With this in mind, our group decided to examine how COVID-19 has affected two different spheres of study related to geography. Drawing inspiration from the traditional split of geography into physical and human, we intend to look at the impact of COVID-19 on climate (physical) and transportation (human). Specifically, we plan on comparing changes in weather and traffic patterns pre-COVID-19 and during the pandemic (a month-by-month comparison between 2019 and 2020). A broad-based analysis will be done followed by a case study on Austin, Texas. We will then take a look at how these two factors are related.

An impact that has not gone unnoticed is the surprising rate at which global climate outcomes have responded to the reduction in industrial, economic and social activities due to the pandemic. For example, New Delhi, the world's most polluted capital, saw clear skies for the first time in years. The PM2.5 concentration there averaged 44.18 (safe level) in March 2020 as opposed to 81.88 in March 2019[2]. It is suggested that countries reducing manufacturing activity, imposing lockdowns, compulsory work-from-home arrangements, reduced air travel, lower vehicular exhausts and dust from construction may have had an impact on reducing greenhouse gas emissions along with PM2.5 particles in the atmosphere[3]. Our research looks into if there indeed has been a statistically significant improvement in global climate outcomes from 2019 to 2020.

On the transport end, COVID-19 has significantly changed our commuting behaviours, due to a combination of factors including mobility restrictions imposed by the government as well as public fear. Both qualitative and quantitative studies have shown a general decrease in transport activity. Global road transport activity saw a decrease of 50% from the 2019 average in March 2020[4]. Even when measures are lifted, transport patterns may not simply revert to what they were. There are many open questions such as whether COVID-19 will spur a switch to greener transport modes and whether reduced transport demand will continue given the revelation that telecommuting was not as difficult as perceived[5]. It is an opportune moment to examine these possibilities, with our focus being road traffic.

Finally, we will bridge the two analyses, to examine if there are any correlations between weather patterns and traffic conditions, specifically in Austin. We will also be noting key events that occurred during the pandemic to observe the effect on both weather and traffic. In this way, we can understand the physical and human world changes in the aftermath of COVID-19 and identify any connections between the two in the context of the city of Austin given loosening restrictions[6].

It is hoped that our analysis will provide some quantitatively-substantiated insight into how our cities might indeed build back better through novel approaches.

2. General Methodology and Experimentation

2.1 Climate

The study of Earth's environment is a complex, dynamic system in perpetual motion, where small changes may result in large, unpredictable interactions at varying spatiotemporal scales (Faghmous and Kumar, 2014). Ganguly and Steinhaeuser (2008) argue that the study of climate change science and climate impacts can

[1] https://www.un.org/en/un-coronavirus-communications-team/launch-report-socio-economic-impacts-covid-19
[2] https://www.channelnewsasia.com/news/asia/covid19-india-lockdown-clear-skies-clean-air-pollution-12662694
[3] https://www.channelnewsasia.com/news/commentary/covid-19-lockdowns-climate-change-coronavirus-12985882
[4] https://www.iea.org/reports/global-energy-review-2020
[5] https://blogs.ei.columbia.edu/2020/07/10/urban-transport-changing-covid-19/
[6] https://www.texastribune.org/2020/09/17/greg-abbott-texas-coronavirus/

greatly benefit from data mining techniques, partly due to the sheer volume (one of the 4 Vs of big data) of climate data available for analysis. Moreover, the weather is constantly evolving and forecasts are highly time-sensitive (Velocity - another of the 4 Vs).

Nevertheless, according to Faghmous and Kumar (2014), there are several challenges in applying big data techniques to climate science, causing its potential impact to lag behind other domains.

1. Climate data is often organised in a spatiotemporal grid, so data in regions of close temporal or spatial proximity tend to be highly correlated, violating any independence assumptions.
2. Climate data is exceedingly heterogeneous with 50 "essential climate variables (ECVs)", each originating from various sources and methods of data collection, similar to the challenge of variety in big data (another of the 4 Vs). It is unlikely that we will be able to satisfactorily reduce all “relevant” data using a composite measure of climate; rather, we will need to contextually examine each variable on its own.
3. Data science has historically focused on statistical and pattern-recognition approaches using attribute-value data, as opposed to knowledge systems. However, climate scientists believe that climate science should be studied using knowledge-based approaches instead of using ‘black box’ models that are not interpretable.

Manogaran and Lopez (2017) adopt a combination of the cumulative sum approach and the bootstrap approach to detect climate change over time. Though such an approach does not directly address the challenges mentioned above, it is a widely-used methodology that enables the application of big data techniques to climate studies. The dataset used in the paper is a panel of raw day-wise climate data from 1979 to 2016 in Tamil Nadu, India, encompassing the following variables: relative humidity, wind, maximum and minimum temperature, and solar energy.

The cumulative sum approach is used to “detect the slow and drastic changes in the mean value of a quantity of interest”, thus allowing analysis of whether the climate has indeed changed on average over time. The method is as follows:

1. Reduce the day-wise climate data (ECVs) to seasonal data.
2. Given n data points $X_1, X_2, \ldots, X_n$ of a climate variable, we compute the average: $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$
3. We compute the cumulative sum value S_i for each of the data points: $S_0 = 0,\ S_i = S_{i-1} + (X_i - \bar{X})$ for $i = 1,\ 2,\ \cdots,\ n$.
4. The change in cumulative sum, S_{diff}, is given by: $S_{diff} = S_{max} - S_{min}$, where S_{max} is the maximum value of S_i and S_{min} is the minimum value of S_i.

The bootstrap approach supplements the cumulative sum approach by computing a confidence level(CL) that the change has indeed occurred within the time frame studied. For iteration 0:

1. Reorder the original n data points $X^0_1, X^0_2, \ldots, X^0_n$ using sampling without replacement. The superscript indicates the iteration of the bootstrap analysis.
2. As above, compute the cumulative sum values.
3. Compute the maximum, minimum and difference in cumulative sum values, $S^0_{max}, S^0_{min}, S^0_{diff}$.
4. Verify if the bootstrap difference S^0_{diff} is less than the original difference S_{diff}.

Repeat steps 1-4, N times. Compute the confidence level: $CL = 100 \times \frac{X}{N}\%$, where N = number of bootstraps performed and X = number of bootstraps for which $S^i_{diff} < S_{diff}$.

The researchers explain that 1000 is usually a satisfactory choice for N, with $n!$ being too large to be feasible. To identify a significant change within the original data, we require a CL > 90% or 95%.

2.2 Traffic

Traffic conditions are a complex phenomenon, and there are a multitude of measurements which are used traditionally to measure traffic, including road capacity, traffic volume, density and annual average daily

traffic (US DOT, 2018). However, these measures are only adept in capturing a portion of the traffic conditions. For example, traffic volume only accounts for the number of vehicles on the road, while ignoring important indicators such as speeds.

Given that single measurements do not sufficiently capture full traffic conditions, the concept of a holistic score has been proposed. A Traffic Performance Score (TPS) was proposed by Cui et. al (2020), which takes into account traffic volume, speed and length of the road segment to measure the traffic performance of a traffic network. The method is as follows:

1. Two traffic stream parameters, namely volume and speed are taken into account. As the three main traffic parameters (volume, speed and density) are all related to each other, only two of them need to be included in the calculation.
2. Length of the road segment is also taken into account and combined with traffic volume to represent vehicle miles of travel (VMT) for the road segments.
3. The various parameters are then combined and the TPS is defined as below:

$$TPS_t = \frac{\sum_{i=1}^{n} V_t^i \cdot Q_t^i \cdot L^i}{\sum_{i=1}^{n} V_f \cdot Q_t^i \cdot L^i} \times 100\%$$

Here, V_t^i represents the speed of each road segment *i* at time *t*, Q_t^i represents the volume of each road segment *i* at time *t*, L^i is the length of *i*-th road segment and V_f is the default free-flow speed. Hence, the resulting TPS will be a value ranging from 0% to 100% where a TPS of 100% will represent the best overall network-wide traffic condition and a TPS of 0% will represent the worst at time *t*.

An alternative score, the Traffic Congestion Score (TCS), was proposed by Lee et. al (2018). This takes into account the average travel speed of the link/road segment v_i and speed limit v_l. Other metrics (such as volume) are not considered since speed is identified as the key concern of drivers.

TCS is calculated using the formula below:

$$TCS_i = \begin{cases} 100\% & v_i \leq 0 \\ 0\% & v_i \geq v_l \\ (1 - \frac{v_i}{v_l}) \times 100\% & otherwise \end{cases}$$

The resultant TCS is a value from 0% to 100%, where 100% indicates total congestion with travel speeds being 0km/h, while 0% indicates no congestion with travel speeds ≥ the speed limit of the road.

2.2.1 Our Proposed Traffic Score (TS)

Our proposed Traffic Score (TS) is a modification of TPS and TCS that takes into account volume, speed and road capacity. These were taken to be the key indicators of traffic condition -- traffic volume is a reliable indicator of congestion, traffic speed is identified as the main concern of user experience and traffic capacity reflects what the ideal traffic conditions would be like.

The Traffic Score of a road (*r*) for a day (*d*) over all time intervals in a day (*t*) can be calculated with the following equation. This is useful for fine-grained analysis on road conditions, in order to find congestion hotspots or underused roads.

$$TS_{rd} = \frac{\sum_{t=1}^{T} V_t^r \cdot Q_t^r}{\sum_{t=1}^{T} SL_t^r \cdot Q_t^r}$$

where T is the number of time intervals in a day, V_t^r is the average traffic speed on the road *r* at the interval *t*, Q_t^r is the traffic volume on the road *r* at the interval *t*, SL_t^r is the free-flow speed of the road *r* at the interval *t*. The **higher** the TS, the **better** the traffic conditions. A TS ≥ 1 indicates good traffic conditions, while a TS < 1 indicates some extent of congestion.

To obtain a macroscopic view of traffic conditions, one might want to obtain a combined Traffic Score across roads in a region. In order to aggregate TS across roads, we take into account the road usage via a weighted average of TS. More popular roads, such as highways, are awarded a higher weight, estimated via the traffic volume to reflect a more accurate traffic condition. The Traffic Score for a day (*d*) over all road can be calculated as follows:

$$TS_d = \sum_{r=1}^{R} \frac{Q_d^r}{Q_d} \cdot TS_d^r$$

where R is the number of roads in the region, Q_d^r is the total volume on the road *r* on day *d*, Q_d is the total volume on day *d*

2.2.2 Road Danger Score

While a Traffic Score provides insight into the health of traffic in terms of important aspects such as speed, volume and road capacity, it would be remiss not to consider safety when evaluating the condition of roads. Indeed, road safety is one facet of journeys that simply cannot be compromised. Hence, to provide a more holistic perspective on traffic conditions, we decided that it is imperative that we generate a Road Danger Score that is indicative of road safety.

Our Road Danger Score is based on previous work done by Archer (2005) and Rumar (1998). Rumar (1998) outlines three main dimensions of traffic safety - risk, exposure and consequences - and proposes a relationship between these three dimensions in a descriptive model as can be seen from the diagrammatic representation by Archer (2005) in Figure 1 below.

As Archer (2005) describes, "each is considered relevant given the fact that changes in any one particular dimension will have an influence on the overall traffic safety situation represented by the total area". As the data we were able to obtain for our case study of Austin was rather different from those utilized by Rumar (1988) and Archer (2005), further discussion of the exact methodology for the Road Danger Score is left for the *Case Study: Austin* section.

Fig 1: "Three dimensions of traffic safety proposed by Rumar (1988)" as illustrated by Archer (2005)

3. Datasets Used

3.1 COVID-19

We will focus on COVID-19 data in the state of Texas for analysis on the change in traffic conditions.

- Texas COVID-19 Data: This website provides data regarding COVID-19 in the state of Texas and will be filtered by the county to utilise only data that is geographically relevant to Austin/Travis County.
- How COVID-19 shut down Texas: This website provides a timeline of key events in the state of Texas and the city of Austin that we will use as part of our analysis.

3.2 Climate

To provide a global view on climate, we use databases that cover a large range of cities:

- Daily global air quality data from the Air Quality Open Data Platform, 2019 Q1-4 and 2020: This dataset includes the minimum, maximum, median and standard deviation for air pollutants over 500

major cities, including PM2.5, PM10 and Ozone. The data reported for each city is averaged across several stations.
- We obtained longitude and latitude information from the geopy open-source library in Python, and they were automatically inferred in Tableau.

3.3 Traffic

All our datasets will be taken from data sources in Austin or the state of Texas. The following data is core to the analysis that we shall be performing:

- Camera Traffic Counts: This dataset provides information regarding vehicular movement at traffic intersections throughout Austin, taken at 15-minute intervals. The key columns of information that we will be using are the columns intersection name, volume (of vehicular movement) and speed average (of vehicles passing through that intersection). These data points will be used to calculate our Traffic Score. At 34.6 million rows and data covering the year 2019 up until the present date, it is sufficiently large and covers long enough duration to be meaningful for our calculations.
- OverPass API: This API provides road information, including speed limits, given lat-lon coordinates. It will be utilized in tandem with the above 'Camera Traffic Counts' dataset to calculate traffic congestion, utilized in the calculation of our Traffic Score.
- Real-Time Traffic Incident Reports: This dataset provides information regarding traffic incidents within the city of Austin. It will be utilized in the generation of a Traffic Score as we seek to generate a score that also takes into account the safety of road usage.
- Traffic Detectors: This dataset provides information on traffic detectors located in Austin, such as coordinate-based location and detector status. It will be utilized together with the 'Camera Traffic Counts' so as to include data only from working detectors. The location information will also be used for plotting our Traffic Scores on a map.

4. Global Analysis

4.1 Climate

While including all 50 ECVs in our analysis would provide a more complete picture of climate conditions, we recognised that it was next to impossible to do so due to the sheer size of climate data. Most other research studies we reviewed focused on 5-6 relevant ECVs. We believe this is because ECVs are often highly correlated yet may be irrelevant to an object of inquiry, as mentioned by Faghmous and Kumar (2004). Hence, we will be focusing on PM2.5 concentrations and temperature in our analysis as they are most likely to be representative of any changes that result from reduced land transport due to the pandemic.

As climate change is very much a global phenomenon that cannot be discussed simply in the limited context of a city, country or region, we aim to conduct a global analysis. Since cities are the biggest contributors to greenhouse gas emissions due to their population density and concentration of economic activities, we will be studying cities around the world to provide a global perspective on climate change by doing a month-to-month comparison between 2019 and 2020.

4.1.1 Preprocessing

Scope of analysis: The dataset we obtained had readings for 8 ECVs such as PM10, humidity, sulfur dioxide in addition to temperature and PM2.5 measurements which we were interested in. Moreover, multiple readings were given for each ECV at a particular location - minimum, maximum, variance and median. There were separate CSV files for each quarter of the year. Thus, we also had to combine the intermediate results together as the final step in preprocessing to save on disk I/O.

These were the steps that we followed to preprocess the climate dataset:

1. Extract PM2.5 and temperature measurements into two separate CSV files. Only extract the relevant columns - city, median ECV reading and ECV type (either PM2.5 or temperature).

2. Process the date columns such that day is removed from the date field. This is to allow for grouping by month in the following step.
3. Collapse the temporary result from the previous steps by grouping by city and date (month) field. The ECV reading field would be aggregated by average.

4.1.2 Data processing

We process the climate dataset obtained using the cumulative sum and bootstrap approach as described above, using Spark. These were the steps for each ECV:

1. Collect distinct observations from the city column to identify the cities that had data for the ECV.
2. For each city:
    a. Extract the relevant data points, i.e. the list of monthly average median ECV readings.
    b. Compute the mean of the monthly average median ECV readings.
    c. Process the data in chronological order. Initialise the cumulative sum to 0. Subtract each monthly average median reading from the mean obtained in (b) and add it to the cumulative sum. Keep track of the maximum and minimum cumulative sum values during this process. The change in the cumulative sum over this period is defined as the difference between the maximum and minimum cumulative sum values.
    d. Since we have selected N=1000 (as suggested by Manogaran and Lopez, 2017) for our bootstrap process, we repeat this 1000 times:
        i. Randomise the order of the monthly average median ECV readings for the city.
        ii. Compute the difference in cumulative sum on the randomly-ordered set.
        iii. If the new difference in cumulative sum value is smaller than the original, add one to the counter of successful bootstraps.
    e. Compute the confidence level, which is the number of successful bootstraps over the number of bootstraps (N=1000).
    f. Use geopy API to get the longitude and latitude for the city, to be used in visualisations.
3. Export a new dataset containing only the confidence level, latitude and longitude of each city.

Upon obtaining our results, we realised that we could only capture how confident we are that a change has occurred, but not what the magnitudes or the directions of the changes were. To find the direction and magnitude of the change, we also took the following steps for each ECV for each city:

1. Obtain a difference value for each month - the difference between the averages in 2020 and 2019 for the corresponding month (i.e. for the month of April in Austin, the difference is average of ECV reading in April 2019 subtracted from the average of ECV reading in April 2020).
2. Obtain the aggregated difference by taking the average across all months computed in step 1. This difference will have a magnitude and direction component.

The data processing for each city can be done on a separate worker node, making this approach highly scalable. Furthermore, the algorithms we use are simple and do not take up much memory, requiring only the set of monthly average median ECV readings to be in memory at one time, which is a dataset whose size is limited by the time period we are considering. The approach can also be easily extended to consider more ECVs simply by tweaking the preprocessing step to extract additional ECVs.

4.2 Traffic

In an ideal situation, we would conduct a broad analysis of the shifts in traffic conditions across the globe. However, we were faced with two main limitations:

- Every city collects data in their own formats and with varying granularity; standardising these disparate datasets would be a mightily time-consuming effort.
- Analysis of traffic must be performed with the specific context of the city being examined in mind. A standardised traffic score may not fully account for the idiosyncrasies between cities.

As such, we perform our traffic analysis specific to the city of Austin, Texas. The reasons for this selection and further details are provided in the subsequent section.

5. Case Study: Austin, Texas

We will be focusing on Austin, Texas for our analysis as it is one of the largest cities in the United States of America and has a similar geographical size as compared to Singapore. Unlike other major US cities like New York City or Washington DC, less research has focused on the traffic and mobility patterns in Austin. Located in Travis County, Austin also has comprehensive as well as granular big data available that we could analyse and gain insights from. Finally, given that Austin has been the "fastest-growing metro"[7] for nine years, it is uniquely positioned to be one of the major new American cities that can be a pace-setter on how to build back better. Hence, Austin is a prime candidate for our case study.

5.1 Climate

Since we have estimated the magnitude, direction and confidence level of a change in temperature and PM2.5 over the 2019-2020 period for all cities in our dataset during the global analysis stage, we simply extract the relevant observations for Austin. In particular, the monthly differences are of interest as they can be plotted against the monthly traffic for Austin so as to investigate the time-series relationships between these variables.

5.2 Traffic

5.2.1 Cleaning

To gain the data that we required from our datasets, some preliminary cleaning had to be performed. This included the removal of data values and columns that were not relevant to our analysis. For example, columns specifying accident details (eg. "status", "address") were removed to provide more abstraction of the accident type and severity during analysis. Data values such as movement around intersections were also removed, as we wanted movement *through* the intersections which would give more accurate traffic speed measurements.

5.2.2 Preprocessing

In order to calculate Traffic Score, the following 3 measurements needed to be prepared: traffic volume, average traffic speeds, and road capacity.

1. **Traffic volume**: this was calculated with a simple sum of heavy and non-heavy traffic volume across all directions through the intersection.
2. **Average traffic speeds**: rather than a simple average, we opted for a weighted average of traffic speeds across all directions through the intersection, weighted on traffic counts. The motivation for this was the observation of noticeable outliers (eg. speeds of 100km/h as compared to the typical speed of <50km/h), which would skew the average traffic speeds.
3. **Road capacity**: this is represented by the free-flow speed of the road, which was estimated via speed limits.

5.2.3 Traffic Score (TS)

The TS was calculated via the proposed steps and equations highlighted in section 2.2.1. Since data was obtained over 15-minute intervals for road intersections throughout the day, the daily TS per intersection was first calculated over all 15-minute intervals in the day using the 3 measurements (volume, speed, capacity) as prepared in the pre-processing step.

However, in order to observe the general trend of traffic conditions, a combined Austin TS would offer a better picture.

[7] https://patch.com/texas/downtownaustin/austin-americas-fastest-growing-city-report

- To aggregate the TS over all intersections in Austin to get a combined TS, we applied our outlined aggregation approach (specified above) to obtain a weighted average of TS across intersections, weighted by road usage.
- This daily combined Austin TS offers a holistic representation of traffic conditions, giving important intersections like highways higher weightage.
- Subsequently, we varied the granularity of the score by aggregating by month to reflect how TS changed across different months of the year.

5.2.4 Road Danger Score

In Austin, we were able to obtain real-time traffic incident reports that were reported to Austin Police and the Travis County Sheriff's Office. While these reports covered a gamut of incidents, we were unable to get more granular information such as the number of injuries and number of accidents that would have allowed us to utilize Rumar's (1988) model in full.

However, Archer (2005) provides some examples of traffic safety risk including health risk in traffic (number of fatalities/injuries per million hours in traffic) and accident risk (number of accidents per million kilometres travelled per person). Building on these categorizations of Archer (2005) and the dimensions of traffic safety proposed by Rumar (2005), our team decided to evaluate the various relevant traffic incident types and categorize them into three discrete severity levels. This can be seen below in Table 1. In addition, to approximate the distances travelled or time spent in traffic (information we did not have access to), we utilized the total volume of traffic across the intersections from the earlier Traffic Score calculations.

SEVERITY 1	SEVERITY 2	SEVERITY 3 (HIGHEST)
● Stalled Vehicle ● Vehicle Fire ● Crash Service ● Traffic Hazard (Debris) ● Collision between Vehicles ● Collision with Private Property ● Traffic Hazards ● Icy Roadways	● Collision with Injury ● Collision (Leaving Scene - Hit and Run) ● Fleet Accident with Injury	● Traffic Fatality ● Fleet Accident with Fatalities ● Automobile and Pedestrian Collision ● Crash Urgent

Table 1: Classification of Incident Severity

Thus, we generated our monthly Road Danger Score as follows, where a lower Road Danger Score is indicative of safer roads: $\text{Road Danger Score} = \frac{\sum_{i=1}^{n} S_i}{\text{total traffic volume}} \times 1000$, where n is the number of traffic incidents and S_i is the severity level of incident i.

6. Results and Discussion of Results

6.1 Climate Results

We plotted our results on two platforms - Python and Tableau. Python allowed for more customisation of how we wished to plot the results, while Tableau allowed for us to make the plots interactive with limited control over the customisations. Plots with Python were plotted with geopy open source library.

We plotted two sets of bubble maps for each ECV (PM2.5 and temperature):

- The first set of maps captures the **average magnitude of change** of the ECV with the **size of the dot**, while the **colour of the dot** represents the **confidence level** (i.e. how confident we are that a change has indeed occurred). With the first map, we wish to capture how much the climate has changed and how confident we are about this change due to COVID-19.

- The second set of maps again captures the **average magnitude of change** of the ECV with the **size of the dot**, but the **colour of the dot** now represents the **direction of change** (i.e. whether the ECV increased or decreased from 2019 to 2020). With the second map, we wish to fill in the gaps of the first map - has the change in climate been positive or negative?

Our results for each ECV will be analysed using the three variables - confidence level, the magnitude of change and direction of change.

- Firstly, for the confidence level, we anticipate that we will be more confident about the change in PM2.5 than temperature as based on our understanding of climate systems, temperature takes a much longer time to change and is linked to many other climate variables as opposed to PM2.5, which can be more directly linked to emissions from manufacturing activities.
- Similarly, for the magnitude of change, we expected that PM2.5 concentrations will have a greater degree of change as opposed to temperature as temperature is likely to be more stable and linked to long-term trends such as global warming instead of being easily affected by temporary changes in greenhouse gas emissions.
- Lastly, for the direction of change, we predicted that both PM2.5 and temperature will show a general decrease from 2019 due to COVID-19.

However, we also recognise that COVID-19 has had different degrees of impact on different cities due to a myriad of reasons - different government policies, degree of spread of COVID-19, recovery rates, etc. We also note that climate cannot be studied at the city-level alone as climate systems of cities all over the world are intertwined and extremely dynamic, such that an increase or decrease in PM2.5 or temperature in one city will affect other cities in its region or even globally.

6.2 Climate Analysis

First, we analyse the set of figures where the size of each dot represents the average magnitude of change for the city and its colour represents the confidence level (CL) of the change.

Fig 2: The colour scale we used to represent the confidence level of the change.

From left to right: $CL < 50$; $50 \leq CL < 80$; $80 \leq CL < 85$; $85 \leq CL < 90$; $90 \leq CL < 95$; $95 \leq CL < 100$.
The blue dots represent dots that are not significant under either a 5% or 10% level of significance.
The orange dots represent dots that are significant if we adopt a 90% or 95% threshold for CL.

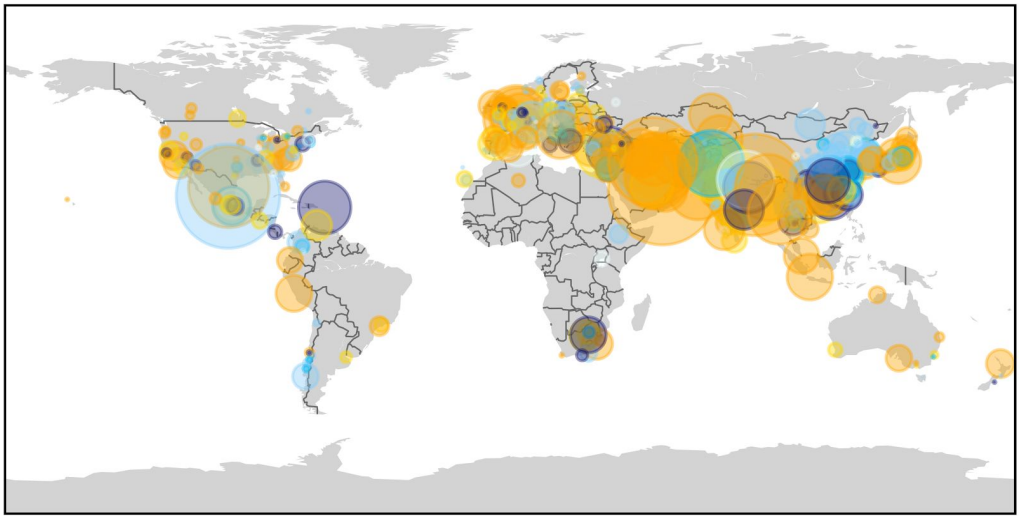

Fig 3a: Magnitude-CL plot for PM2.5

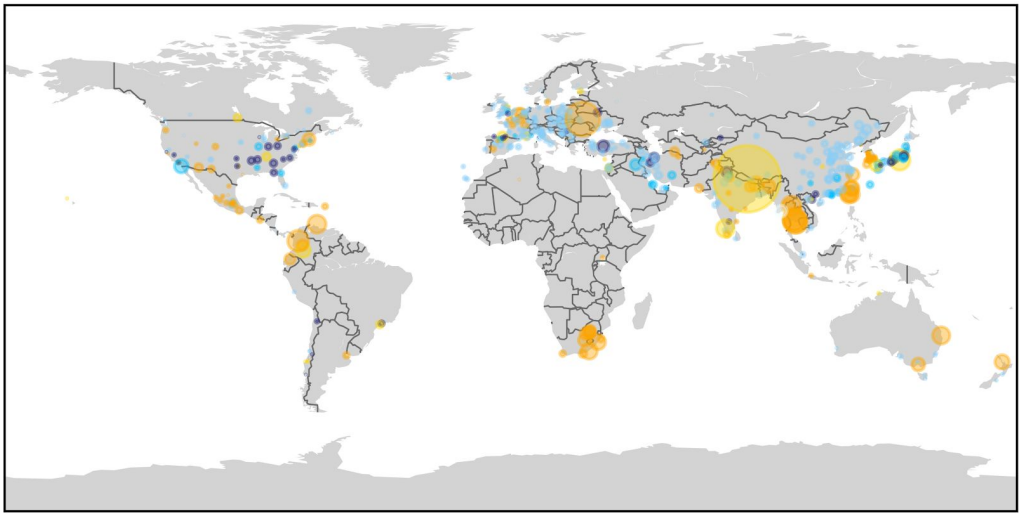

Fig 3b: Magnitude-CL plot for temperature

Our results appear quite promising! If COVID-19 has indeed had a significant impact on the climate, we would expect to see a number of orange dots in both the PM2.5 and temperature plots.

For PM2.5, there is a large number of orange dots across Asia, Australia, Europe and the US, with those in Asia being especially large in magnitude. This represents a large and significant effect of the COVID-19 pandemic on PM2.5 emissions in various cities in 2019-2020. Indeed, this is consistent with our expectations as we expect PM2.5 emissions to be lower this year in places like New Delhi, which we mentioned in our introduction. As for temperature, though the dots appear much smaller than for PM2.5 (which is also consistent with our hypothesis), and there are fewer orange dots, there are still a number of significant changes in some cities, such as in Asia, Africa and South America.

Next, we look at the set of figures where the size of each dot represents the average magnitude of change for the city and its colour represents the average direction of the change. Red dots represent a positive change in the ECV on average (i.e. its average values increased between 2019 and 2020) while blue dots represent a negative change.

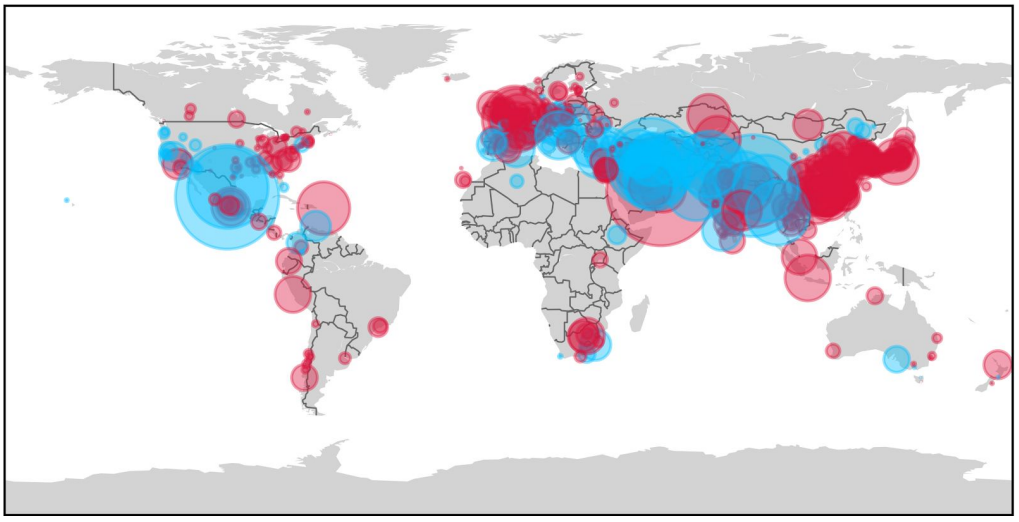

Fig 4a: Magnitude-direction plot for PM2.5

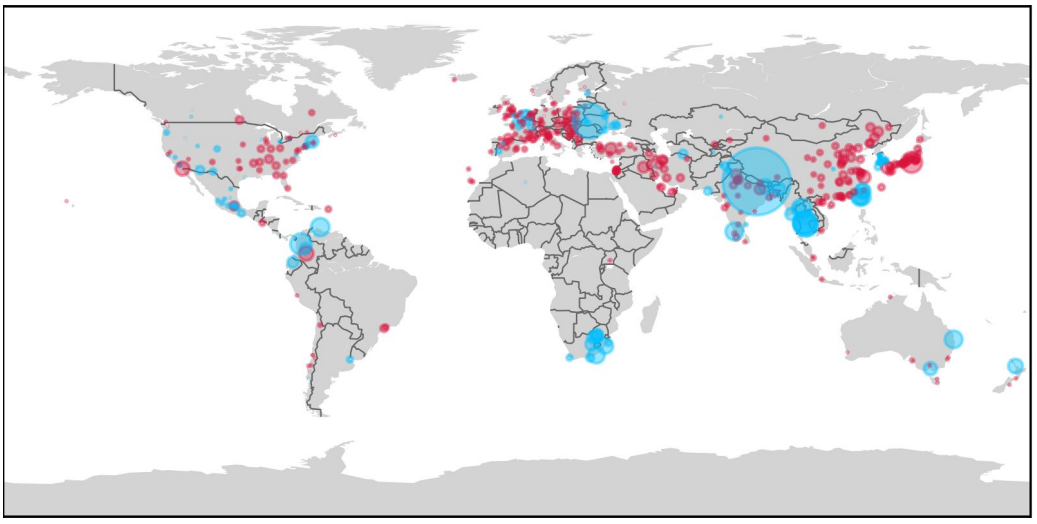

Fig 4b: Magnitude-direction plot for temperature

Our results appear to be mixed. For PM2.5, there seem to be many large blue dots especially in Asia and the Middle East, although there are also many large red dots in East Asia and Europe. This means that some of the cities with large and significant changes we observed in the previous plot had, in fact, experienced *positive* changes in PM2.5 between 2019-2020 instead of negative changes like we had anticipated due to COVID-19. As before, the temperature results appear much more muted, with only a few large blue dots globally - in many of the same locations as the large orange dots in the previous plot! - and mostly small red dots elsewhere.

While the results may seem to be mixed, nonetheless we are able to conclude that the COVID-19 pandemic has had a significant effect on the two ECVs. This is because in the absence of COVID-19, we would have expected most, if not all, of the cities to have experienced *positive* changes in both ECVs over the time period due to global warming. Interestingly, we note that most of these blue dots are primarily concentrated in areas that have been hit harder by COVID-19, such as the US, while red dots are found in areas which recovered relatively faster (barring recent developments in November), such as East Asia (China, South Korea etc). This likely speaks to the duration of COVID-19's impact on economic and human activities in these cities, with harder-hit cities possibly having reduced their emissions more and thus having experienced more reductions in temperature and PM2.5.

These results also confirmed our intuition that PM2.5 would be more responsive than temperature to changes in emissions due to COVID-19, in both magnitude and confidence level. We reproduced the above plots on Tableau to create an interactive visualisation, which you can explore at the link provided.

6.3 Traffic Results and Analysis

Our results were plotted in 2 different ways: firstly, as trends across different time periods in order for the reader to understand how Traffic Scores and Road Danger Scores varied with time and other factors like daily COVID-19 cases; secondly, in a graphical representation for the reader to quickly gain an understanding of our results over a geographical area using Tableau.

The following trends reflect some of our key findings:

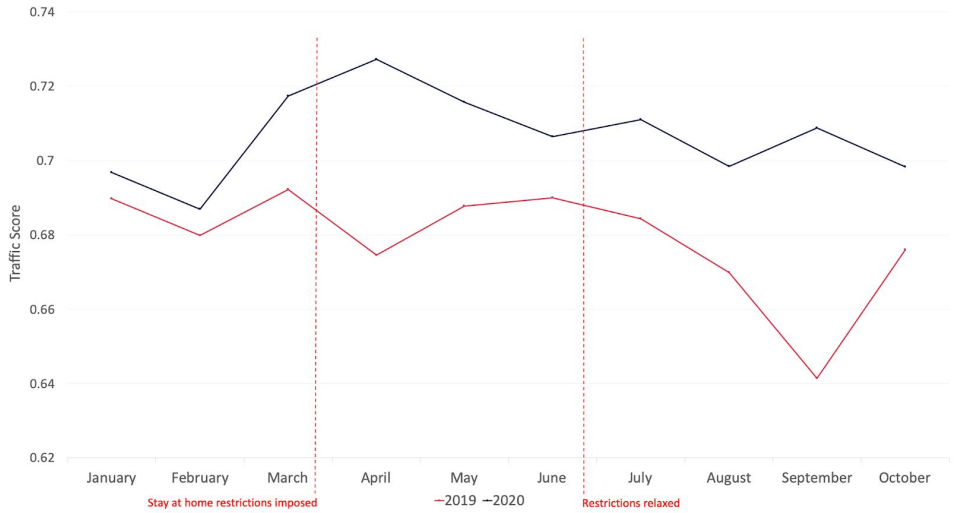

Fig 5: Comparison of Austin Traffic Scores across 2019 and 2020 (January - October)

Our first graph reflects a month-by-month comparison of Austin Traffic Scores for the same period of January to October in 2019 and 2020. As can be seen from the graph, the Traffic Scores for Austin in 2020 were higher for all months as compared to 2019, indicating that traffic conditions were better in all months of 2020. We can also see the impact of COVID-19 restrictions on the Traffic Scores as even though the scores for January and February were similar for both 2019 and 2020, the scores diverged significantly from March onwards as stay-at-home restrictions were imposed and an increasing number of people were working from home along with a general decrease in travel and activities.

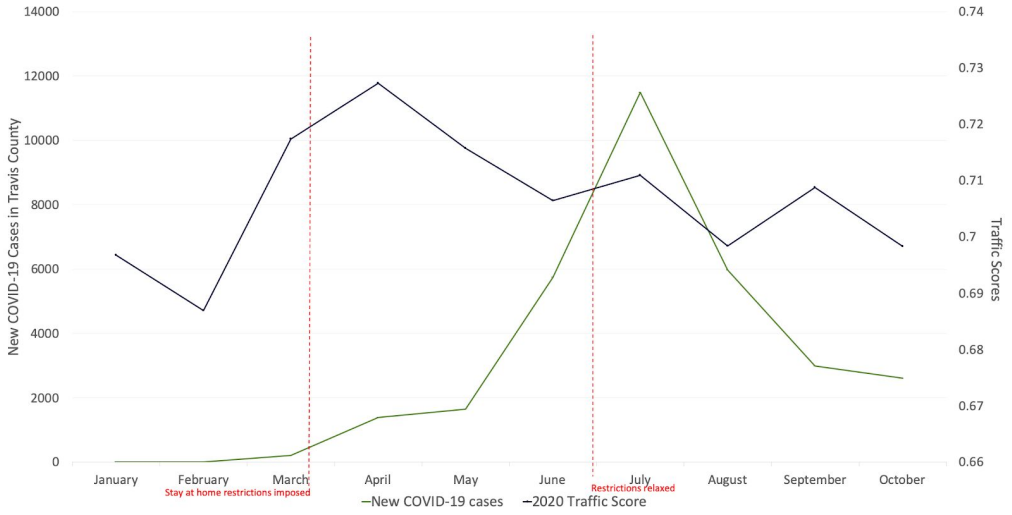

Fig 6: Comparison of Austin Traffic Scores and new COVID-19 cases in 2020 (January - October)

Our next graph plots the Traffic Scores for Austin alongside the new COVID-19 cases in Travis County between January and October this year to see how our Traffic Score responded to the changes in COVID-19 cases. As we can see from the graph, the increase in COVID-19 cases between February and April corresponded to an improvement in Traffic Scores, possibly due to the newly imposed restrictions. Subsequently, we can see that May onwards there was a slight but general decrease in the Traffic Score, even as new COVID-19 cases were rising and peaked. We posit that this trend may be attributed to the increased preference for private transport over public transport due to the fear of coming in contact with other people. This has been substantiated by research into consumer sentiments about the perceived risks of using public transport during the pandemic. In a survey conducted by IBM[8], 20% of people who regularly used buses, subways or trains said

8 https://newsroom.ibm.com/2020-05-01-IBM-Study-COVID-19-Is-Significantly-Altering-U-S-Consumer-Behavior-and-Plans-Post-Crisis

they no longer would, while 17% of people indicated they would use their car more due to COVID-19. This would have prompted an increase in the utilisation of private transportation, leading to a slight deterioration of traffic conditions over the months resulting in lower Traffic Scores from May onwards.

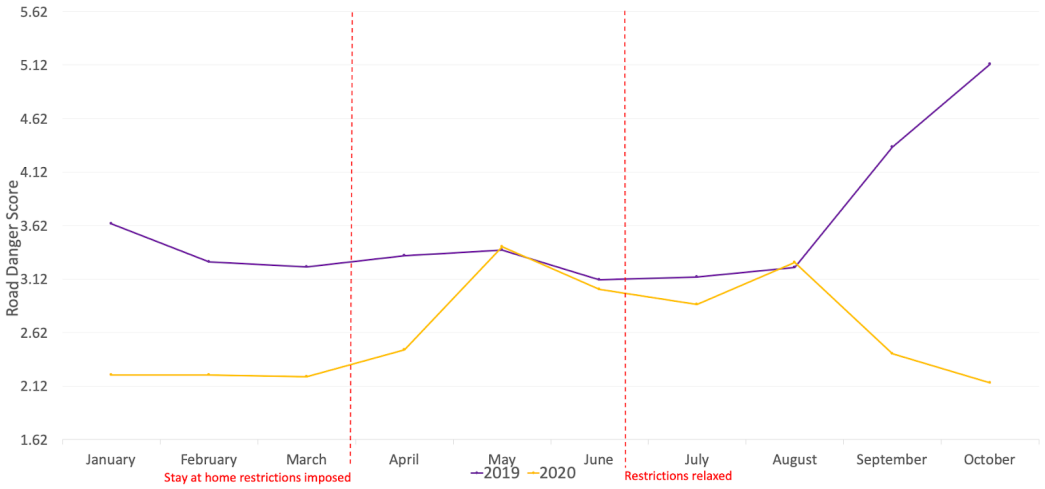

Fig 7: Comparison of Austin Road Danger Scores across 2019 and 2020 (January - October)

Our next graph does a month-by-month comparison of Road Danger Scores for Austin between the months of January and October for 2019 and 2020. Our Road Danger Score measures how dangerous roads are in general, so a lower Road Danger Score implies safer roads. It can be seen that 2020 was generally safer as compared to 2019 for the same period. The only place where we observe a similar Road Danger Score is between the month of May to August. This was the time period when Travis County was experiencing a surge in COVID-19 cases which possibly led to an increase in the use of private transportation as mentioned above, resulting in accidents happening more often, increasing the Road Danger Score.

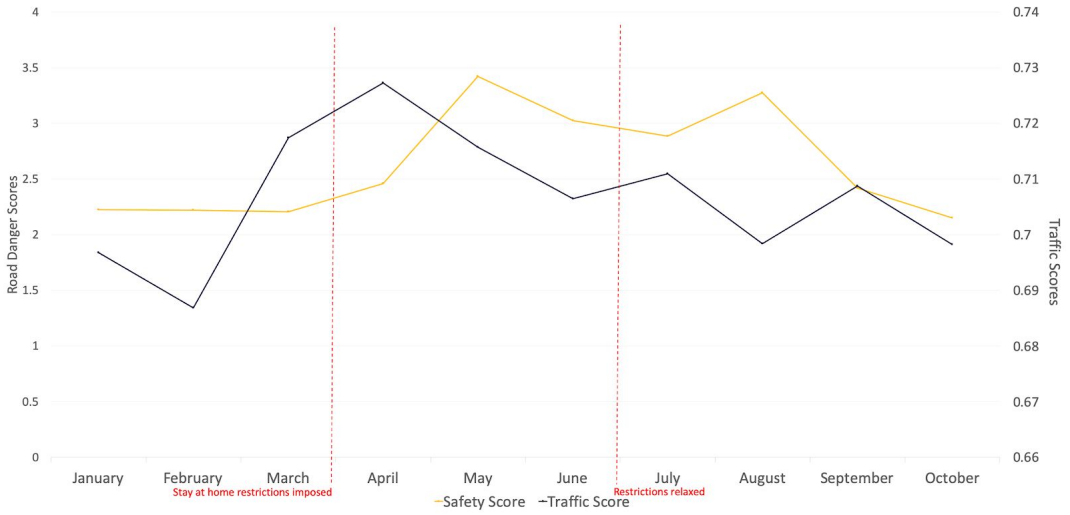

Fig 8: Comparison of Road Danger Scores and Traffic Scores in Austin in 2020 (January - October)

Analysing both our Road Danger Scores and Traffic Scores together for Austin between the month of January to October for this year gives us a very interesting trend. As can be seen from the graph, for certain months like June-September, whenever our traffic scores increased, our danger scores decreased and vice versa. This provides some confirmation for a very intuitive hypothesis that as there is a worsening of traffic conditions (leading to a lower traffic score), it leads to roads being more unsafe (a higher danger score). However, the analysis is more complex than this. This is because while we may expect an increase in accidents and hence higher Road Danger Score due to worsening traffic conditions, accidents may also decrease due to a reduction in speeds of cars because of the increased congestion. Retallack and Ostendorf (2019) also postulate that accident occurrence may be a function of both the speed and volume of cars on the road. That is, at low

traffic volumes accidents occur frequently as a result of high speeds, while at high volumes, accident rates are higher due to more cars being on the road.

6.4 Climate and Traffic in Austin

For our combined analysis of climate data and traffic data, we decided to look at it by analysing PM2.5 values and our calculated Traffic Scores for Austin from January to October 2020. We chose PM2.5 values over temperature because intuitively, PM2.5 is likely to be more sensitive to changes in traffic conditions, as vehicular exhaust contributes directly to PM2.5, as argued by Askariyeh et. al (2020). Furthermore, from our global analysis earlier, we saw that PM2.5 values are more significantly affected by COVID-19 over the period of 2019-2020 than temperature, suggesting that it is more responsive.

By analysing the graph of the combined results, we found a negative correlation between PM2.5 values and Traffic Scores. That is, PM2.5 scores increase with worsening traffic conditions and drop with improved traffic conditions. There are two main factors that contributed to the changes in traffic conditions and consequently Traffic Scores and PM2.5 values:

- The general increase in Traffic Scores in the first few months of 2020 can be attributed to the overall fall in transport mobility due to the imposition of stay-at-home orders and an increasing number of people working from home. The general drop in traffic would explain the drop in PM2.5 values caused by a reduction in pollution.
- Subsequently, there was a slight but continuous fall in Traffic Scores caused by an increased preference for private transport over public transport due to COVID-19. This would mean that there are more cars on the roads, attributing to an increase in PM2.5 values.

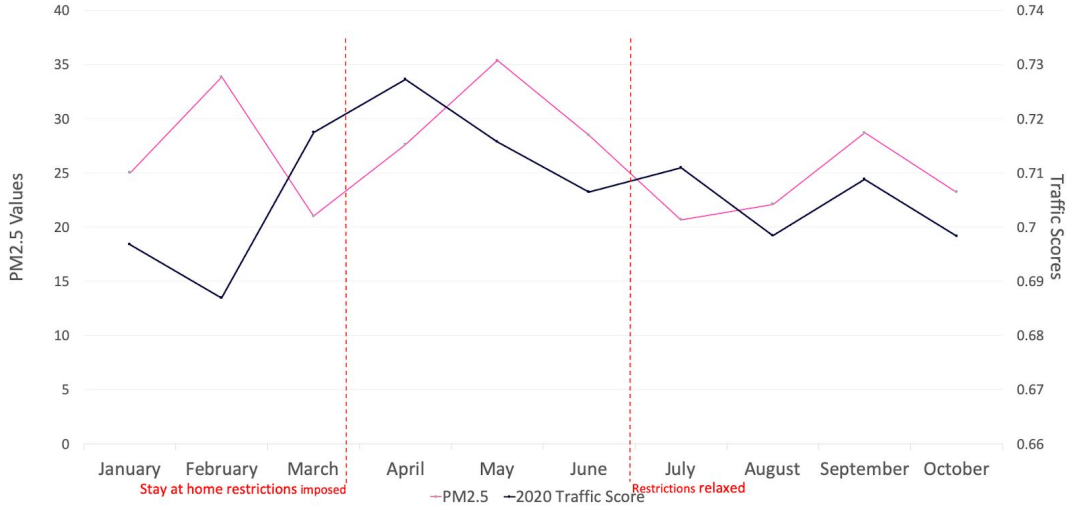

Fig 9: Traffic Scores and PM2.5 values for Austin between January-October 2020

6.5 Interactive Visualisation - Tableau

In order to display our findings, we created two interactive data dashboards using Tableau, a data visualisation tool. These dashboards contain data at a higher level of granularity that would be best viewed with user interaction. For example, in our interactive maps for climate data, the user can hover over cities to obtain the confidence level, the direction of change and magnitude of change. Our dashboards include:

Climate Dashboard
(https://public.tableau.com/profile/chelsea2643#!/vizhome/cs4225wb-final/Climatedashboard?publish=yes)

- World temperature differences and CL for all cities between 2019 and 2020
- World PM2.5 differences and CL for all cities between 2019 and 2020

Traffic Dashboard
(https://public.tableau.com/profile/chelsea2643#!/vizhome/cs4225wb-final/Trafficdashboard?publish=yes)

- Traffic Score for each intersection in Austin for each day (date slider available to change date)
- Traffic Score per day across 2019 and 2020 (January - October)
- Monthly Road Danger Score difference across 2019 and 2020 (January - October)

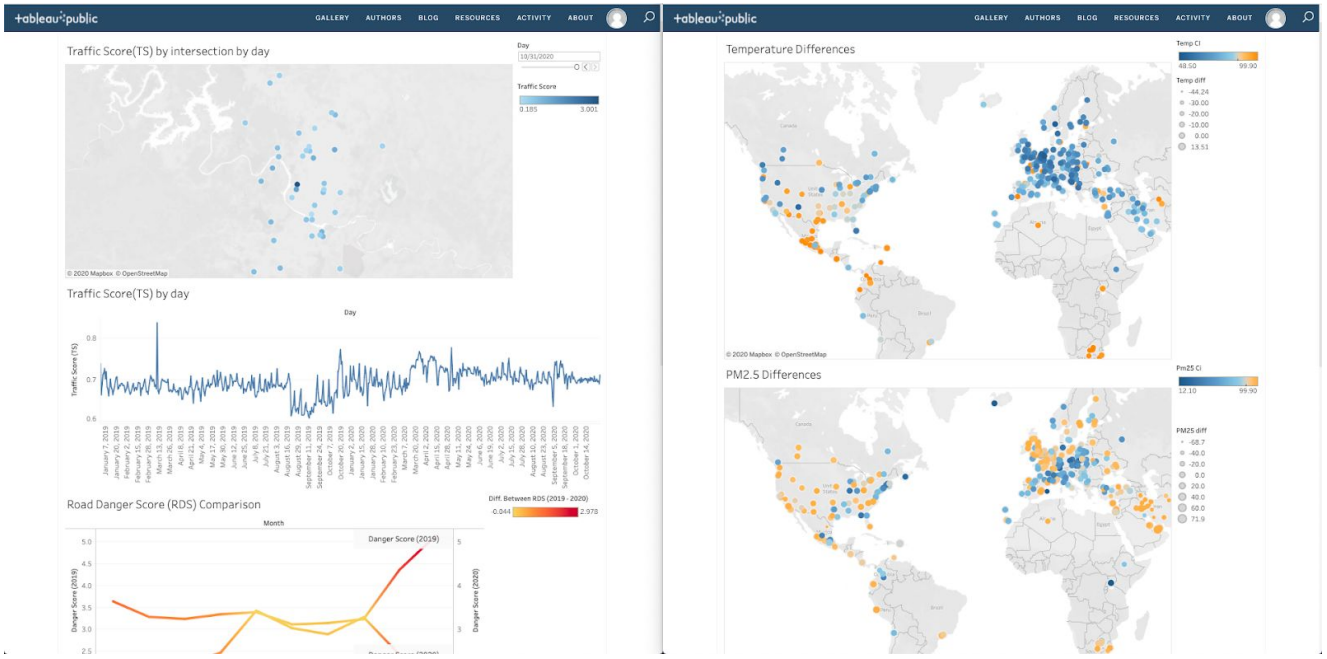

Fig 10: Tableau dashboards for traffic (left) and climate (right)

7. Application: Moving Forward (Literally)

For a start, our proposed Traffic Score and Road Danger Scores can be applied to any city or region, to produce a real-time Traffic Score dashboard to monitor a city's traffic conditions more holistically, similar to the Tableau dashboard we have implemented thus far. This would be particularly useful in all stages of road infrastructure management by urban planners, especially given the more complete view that Traffic Scores and Road Danger Scores can provide compared to simple measures of congestion. For example, it allows for insights into where transport infrastructure needs to be improved to support increasing demand, while at the same time helping to identify roads with heavy traffic loads so that they receive more maintenance attention. It could also be expanded to provide a big-picture view of traffic conditions in multiple cities, such as with a choropleth map. Nonetheless, we note that these scores need to be tailor-made to the individual contexts of each city and are subject to the relevant data being available.

However, our project findings can serve as the basis for more drastic change too. Based on our graphs, we noted that there is a negative correlation between Traffic Scores and PM2.5 levels, i.e. while Traffic Scores improve, PM2.5 levels decrease. Furthermore, there is also a positive correlation between our Road Danger Scores and PM2.5 levels: road danger decreases occur in tandem with PM2.5 level decreases (based on the relation between TS and RDS). Although our sample size is small, these preliminary results we observed are encouraging. Should we continue this analysis over subsequent months, we can surely obtain even more significant results and a better estimate of the degree of correlation between PM2.5 and Traffic Scores. Therein lies an opportunity to solve not just two but three problems in one go: improving mobility, road safety *and* pollution levels.

One possibility of doing so is through encouraging the use of shared micromobility services. Shared micromobility is "an innovative transportation strategy that enables users to have short-term access to a mode of transportation on an as-needed basis" (Shaheen & Cohen, 2019) for short-distance travel. These can take the form of docked bicycle-sharing or dockless scooter sharing, amongst many others.

We suggested earlier that it is likely people opted for private transport as an alternative to using public transport during the pandemic in order to reduce the possibility of coming into contact with a potential carrier. Such behaviour seems to have been particularly pronounced at the height of stay-at-home measures. However, private transport need not be the only safe, convenient mode of transport at this time. Shared micromobility modes are utilised outdoors where the "virus is much, much less likely to spread" (Koerth,

2020), affirming these modes' safety. Thus, shared micromobility provides all the benefits that private transport modes do, and at a fraction of the price, also making it a more inclusive form of transport.

In the long term, shared micromobility needs to be encouraged as a part of "multimodal micromobility and transit trips to replace longer car trips" (McQueen et. al, 2020). Combining different forms of non-car based modes of transport to replace increasingly longer distances of commute would mean reducing the burden on current traffic infrastructure and improving mobility in general. Furthermore, "micromobility's potential to decrease GHG [greenhouse gas] emissions through automobile trip substitution is promising, especially for e-scooters" but is "elusive in many cities, especially if travelers lack preference for this behavior" (McQueen et. al, 2020). This speaks to an urgent need to promote such modes of transportation in order to truly realise the unbridled potential it has to offer.

7.1 Contextualising to Singapore

As mentioned, the city of Austin was selected due to its similarities to Singapore -- both cities are capital cities, have similar geographical size, experience similar population growth, and have well-connected traffic and public transport systems. Thus, it is likely that our findings in Austin will be applicable to Singapore.

Singapore also has an environment where shared micromobility programs can be easily launched given the country's relatively small size. With strong institutional support, shared micromobility players and researchers in the intersection of climate and traffic can use Singapore as a testbed for R&D in this area. These programs themselves will produce big data that can be harnessed to further optimize transport and achieve climate goals, serving as a model for others in the region to follow. For all the loss that the pandemic has wrought, the least we could do is not to squander this opportunity to do things better.

8. Big Data Systems Architecture & Optimisations

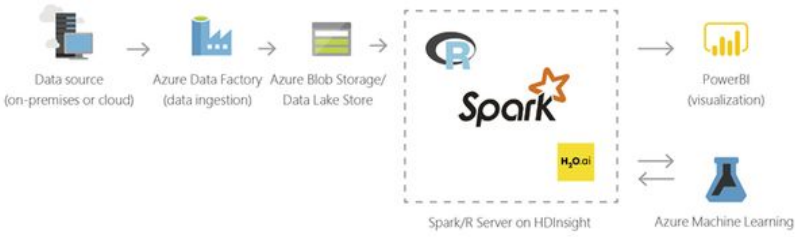

Fig 11: Azure HDInsight architecture (source: Microsoft documentation on HDInsight)

Azure's managed open-source data analytics service, HDInsight, enabled us to streamline our data analysis pipeline by taking care of complicated infrastructure matters. With HDInsight, we were able to process large amounts of data more quickly than we would have otherwise been able to on our local machines which have significantly limited memory capacity.

Our climate and traffic data, in the form of CSV files, were stored in Azure Blob Storage (Fig 12) which utilizes an HDFS interface allowing our tool of choice, Spark, to interact with it. Data access was efficient as the compute clusters and data storage were located in close proximity to one another in the same Southeast Asia data centre and was connected by a high-speed internal network.

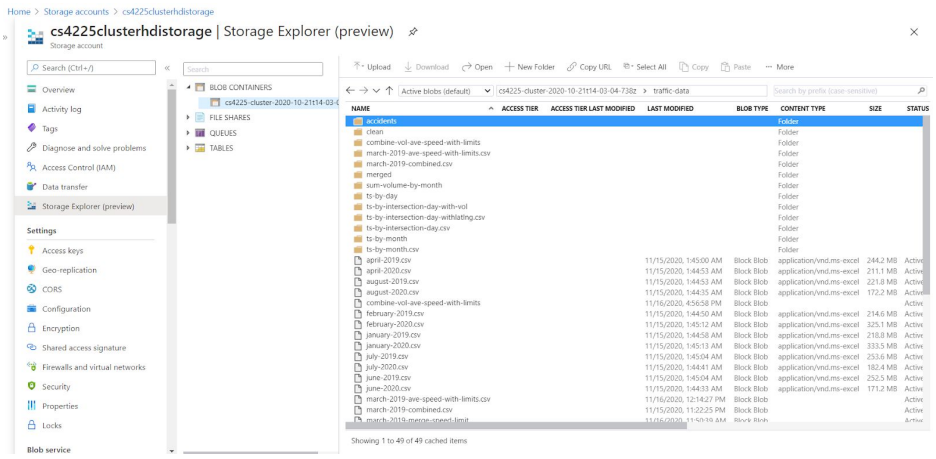

Fig 12: Azure Blob Storage on Azure Console

Crucially, with the decoupling of these components (storage and computation), our use of credits are optimised. As seen in Fig 14 below, this meant that we could add and remove compute worker nodes depending on usage, without losing any data. In the future, if we were to perform computations on larger sizes of data, we simply need to add additional worker nodes to the cluster (i.e. scale out). However, if there is a situation where scaling out is insufficient and the data cannot be stored only in-memory, Azure provides compute shapes with larger memory that can be utilized (i.e. scale up). The power of the cloud is clearly demonstrated here.

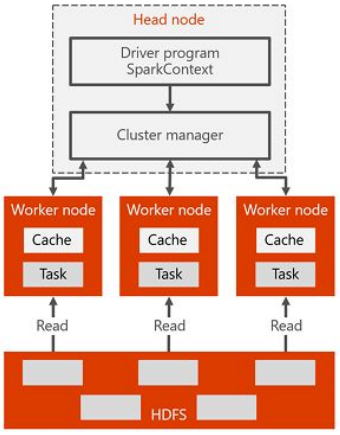

Fig 13: HDInsight Cluster for Spark (source: HDInsight documentation)

Our team chose Spark primarily because in-memory processing was a huge advantage over Hadoop MapReduce. Instead of having intermediate results be written to disk which would have resulted in significant disk I/O and network usage, we took advantage of Spark's transformation and action operations which allowed us to also save on DRAM usage by relying on lineage, without any data loss as the original data was still in the HDFS (in our case, Azure Blob Storage). This is illustrated in Fig 15. In addition, our climate analysis involved a significant amount of iterative computation to calculate the cumulative sum which also plays to Spark's strengths.

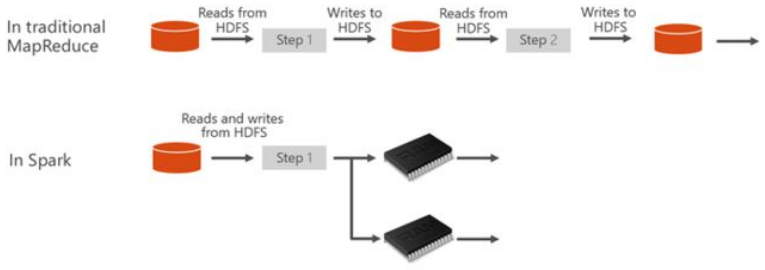

Fig 14: Spark vs MapReduce (source: HDInsight documentation)

With the size of the prepared datasets being around 20GB and thus well below the 32GB RAM that each of our nodes housed, it also meant that there would be no spill to disk and thus lower disk I/O.

Furthermore, we were operating entirely with structured data in both the climate and traffic contexts and with Spark's DataFrames providing us automated schema inference, we could work on the data easily instead of spending time creating object structures to map the data to. Moreover, the computations we were performing were primarily iterative algorithms and Spark lent itself well to this task. Spark's speedy processing was further boosted by Spark's optimization of code execution through logical improvements (Catalyst Optimizer) and its efficient use of CPU and memory (Project Tungsten).

Last but not the least, Spark's API provided us with the flexibility to program in our language of choice and the Spark shell's interactive Read-Print-Evaluate-Loop (REPL) meant that we could try out new code with ease. HDInsight even came with built-in Jupyter notebooks meaning that we could even potentially use PySpark to interactively work with the data and create visualizations.

9. Problems Encountered & Lessons Learnt

Despite the advantages conferred by Spark and using Big data systems to process the large amounts of data we were working with, we faced a few difficulties along the way, particularly in the data sourcing and understanding of metadata.

9.1 Problems with Data Sourcing

Finding relevant big data sources was a time-consuming process. While IoT sensors and other data collection points have become ubiquitous in daily life today, access to that data is not as easily accessible for the following main reasons:

1. Data is oftentimes collected by private entities meaning that it would require monetary investment to use. This was the situation we faced when considering the use of TomTom navigation data.
2. Data collected by government organisations are often confidential and needs specific authorisation levels to access.
3. Raw data is also often already processed to some degree. This is particularly so in scenarios where the data involved can be used to identify individuals and obtain other sensitive information, as such the data is aggregated for obfuscation. Additionally, as the data we were using for understanding traffic in Austin was from a public institution, they were especially judicious in the types and amount of data released for use.
4. For climate data, it is difficult to obtain global datasets that contain the same set of variables measured to the same degree of precision from the same time period. Some large datasets from reputable sources have the data for more recent years behind a paywall, or are no longer actively maintained[9]. If we would like to conduct larger-scale analysis going further back in time, we might need to obtain country- or city-specific data from disparate sources for analysis.

As described in the methodology for the Road Danger Score calculation, we had occasionally had to use whatever data that was available to us as a proxy for the data we actually required. Having access to the actual data required would provide far more accurate impressions of road safety in Austin.

9.2 Problems with Poor Metadata

However, even with access to all the data we needed, it would not be of help if the metadata provided is not complete and informative. When working with our datasets, we encountered many instances of insufficient or poorly written metadata. This meant that we had to make inferences on what some columns and its entries meant based on contextual understanding. For example, one entry of incident type in the Austin Traffic

[9] https://climatedataguide.ucar.edu/climate-data/global-temperature-data-sets-overview-comparison-table

Incidents dataset was “COLLISN/ LVNG SCN”. Without sufficient metadata, we had to take an educated guess that “LVNG SCN” was short-form for “leaving scene”, thus describing hit-and-run situations. In general, inadequate metadata will lead to non-domain experts finding it tedious to work on the datasets provided to them and possibly drawing flawed inferences.

9.3 Algorithmic Design Learnings

Finally, on the note of performance of our algorithmic design we learnt the following lessons:

- While not the focus of our project, if we wanted to perform a similar detailed analysis of traffic or climate conditions, for example in real-time, across an even greater number of cities or a greater number of variables in each domain, our algorithmic design ensures that the addition of more worker nodes (e.g. a new node per city) can handle such a load. This allows for our project to be easily scaled.
- When working with big data, especially on the cloud where data storage and compute are decoupled, minimising network I/O is crucial. To achieve this, bulk sending and receiving of the data files from Azure Blob Storage was key. Once the data files were retrieved, as all our processing was done in-memory using Spark; this meant reducing (or rather eliminating) the amount of hard disk I/O.

10. Project Summary

With COVID-19 proving to be a generation-defining challenge, the question on everyone's mind is: how do we respond? Are we merely going back to the way things were, or are we going to build back better? In our project, we aimed to see the impact of COVID-19 through two different spheres that were inspired by geography: on the physical world and the human world. In particular, we looked at climate and traffic patterns respectively. Intuitively, one might be led to believe that if traffic decreased, climate conditions might improve. However, only a proper, data-backed examination could truly ascertain if such an intuition proved true. More generally, we believe that the results of our analysis may provide some insight on the way human behaviour has changed due to the exigencies of the pandemic.

For our climate analysis, we delved deep into the study of two variables: temperature and PM2.5 values. Firstly, our calculations indicated we would be able to more confidently suggest that changes in PM2.5 levels had occurred as opposed to changes in temperature. On the whole, it was observed that PM2.5 levels had decreased during the height of the pandemic, as compared to the previous year (2019). On the other hand, it was harder to ascertain the direction of temperature change during this period as temperature changes generally take a longer period of time for a trend to be detected, and we would need to contend with long-term trends of global warming. Thus, a more comprehensive analysis using more historical data may be a potential expansion: it would only help prove the validity of our hypothesis, but also expand the sample size for establishing a relationship between PM2.5 data and Traffic Scores. Of course, this analysis can also be continued into the future as the pandemic continues to affect our lives; we could draw more linkages between the recovery trajectories of specific countries and their PM2.5 and temperature measurements.

When it came to traffic, we zeroed in on the city of Austin, the fastest growing city in the US and one that has a lot in common with Singapore in terms of demographic mix and the cosmopolitan nature of activities occurring in the area. To make our traffic conditions analysis holistic, we analysed traffic conditions such as speed and volume using a Traffic Score and complemented it with a separate Road Danger Score that accounted for the equally important aspect of safety conditions. It is interesting to note that both Traffic Scores and Road Danger Scores in 2020 were better or comparable to 2019. Traffic Scores increased particularly sharply in March and April 2020, likely due to a shift toward telecommuting. However, Traffic Scores began declining after that, possibly due to looser adherence to stay-at-home orders and people eschewing public transport in favour of private transport to move about. Predictably, as Traffic Scores fell, Road Danger Scores increased.

To bring the two domains of climate and traffic together, we decided to focus on whether there were correlations between PM2.5 and Traffic Scores as we believe that the short timeframe of analysis meant that it was unlikely that we could discover any meaningful relations between temperature and Traffic Scores. The identification of a negative correlation between PM2.5 levels and Traffic Scores led us to propose shared micromobility as one potential area where our nascent findings could be of use, with special focus on the contextualization of its potential to Singapore.

The pandemic is still ongoing. Some countries are facing caseloads in their second (or even third) wave that are far higher than in the first. Accordingly, the government and citizen responses continue to change by the minute. Our analysis is only a point in time snapshot of the changes that have been observed but are no indication of the permanency of these changes moving forward. However, in such a disruptive moment there lies a crucial opportunity to build back better. Mass adoption of shared micromobility services is one of many enablers of such change. The work done in this project is but a launchpad for developing more far-reaching and comprehensive ways of revolutionising transport and stopping (or even reversing) our deteriorating climatic conditions.

11. References

Cui, Z., Zhu, M., Wang, S., Wang, P., Zhou, Y., Cao, Q., . . . Wang, Y. (2020, July 01). Traffic Performance Score for Measuring the Impact of COVID-19 on Urban Mobility. Retrieved October 06, 2020, from https://arxiv.org/abs/2007.00648

Lee, Jiwan & Hong, Bonghee. (2014). Congestion Score Computation of Big Traffic Data. 10.1109/BDCloud.2014.64. Retrieved October 06, 2020, from https://www.researchgate.net/publication/270894492_Congestion_Score_Computation_of_Big_Traffic_Data

U.S. Department of Transportation (DOT), Federal Highway Administration, Office of Highway Policy Information. (2018). *Traffic data computation method: pocket guide*.

IEA. (2020). (rep.). *Global Energy Review 2020*. Paris. Retrieved October 06, 2020, from https://www.iea.org/reports/global-energy-review-2020

Manogaran, G. & Lopez, D. (2017). Spatial cumulative sum algorithm with big data analytics for climate change detection. *Computers & Electrical Engineering*. doi:10.1016/j.compeleceng.2017.04.006.

Ganguly, A. R., & Steinhaeuser, K. (2008). Data Mining for Climate Change and Impacts. *2008 IEEE International Conference on Data Mining Workshops*. doi:10.1109/icdmw.2008.30

Faghmous, J. H., & Kumar, V. (2014). A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science. *Big Data*, 2(3), 155-163. doi:10.1089/big.2014.0026

Archer, J. (2005). Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling : a study of urban and suburban intersections (PhD dissertation). KTH, Stockholm. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-143

Rumar, K. (1988) Collective risk but individual safety, Ergonomics, 31:4, 507-518, doi: 10.1080/00140138808966695

Shaheen, S., & Cohen, A. (2019). Shared micromobility policy toolkit: Docked and dockless bike and scooter sharing. UC Berkeley Transportation Sustainability Research Center. https://escholarship.org/uc/item/00k897b5

Koerth, M. (2020, October 19). What A Summer Of COVID-19 Taught Scientists About Indoor vs. Outdoor Transmission. Retrieved from https://fivethirtyeight.com/features/what-a-summer-of-covid-19-taught-scientists-about-indoor-vs-outdoor-transmission/

McQueen, M., Abou-Zeid, G., MacArthur, J., & Clifton, K. (2020). Transportation Transformation: Is Micromobility Making a Macro Impact on Sustainability? Journal of Planning Literature. doi:10.1177/0885412220972696

Retallack, A. E., & Ostendorf, B. (2019). Current Understanding of the Effects of Congestion on Traffic Accidents [Abstract]. *International Journal of Environmental Research and Public Health, 16*(18), 3400. doi:10.3390/ijerph16183400

Askariyeh, M. H., Zietsman, J., & Autenrieth, R. (2020). Traffic contribution to PM2.5 increment in the near-road environment. Atmospheric Environment, 224, 117113. doi:10.1016/j.atmosenv.2019.117113