Abdulrahman Alhaidari
Final Project Report- INFSCI 2415

# Workplace mobility changes in three U.S. states between 2020 and 2022 compared to ML predictions

In early 2020, people started working remotely due to the COVID-19 pandemic. As the world is recovering, at the time of this writing in 2022, some people work in a hybrid mode while others came back in-person to workplaces. In this project, the interest to see how the recovery progresses on people going to workplaces in the largest, smallest U.S. populated states compared to Pennsylvania for five consecutive months between 03/01 to 08/1 of each year in 2020, 2021, and 2022. Also, the goal is to see how Machine Learning (ML) models can predict the recovery trend for the year 2022 based on the previous two years' data. Then, the best-performed model will be chosen for the visualization comparison.

Google mobility data [1] was used for 2020, 2021, and 2022. The data has mobility changes for six categories, including workplace mobility, for most countries. Each country divided into regions or states, and each states has multiple counties and cities belongs to those counties. The dataset is publicly available and variable in terms of the data volume based on the data acquisition date. For instance, if I have two versions, one was downloaded a month ago, and a second version was acquired today, the latter will have more data considering the continuous data collection by Google, as it produces a daily report for all countries in all six categories. Even though data reports those six categories of people's movement, some countries have inconsistent and missing data. Overall, the data shows people's activities change over time from the baseline, which was 0% before COVID-19.

Therefore, the first step was to understand the data, as understanding the data before preprocessing enables flexibility when visualizing. To begin with, as the data has various issues that could affect the visualization, it is a necessary step to ensure we consider only relevant quality data. For example, some instances have missing values in multiple features, which could adversely affect the produced visualization, while others are irrelevant to work mobility and cannot be used as it is human unreadable and designated, for instance, to identify a specific location. Again, the interest in work mobility patterns during and post COVID-19.

In this context, I tested multiple ML models results and compared it with the actual data to see which produces the best result, example of which are Neural Networks, linear regression, KNN regression, and random forest regression. For each training variety of tuning parameters were tested. From candidate models, the best model produces the highest performance among others in terms of R^2, MSE, and RMSE for the actual mobility for 2022 to be included for visualization for this project. Again, the main aim is to see if ML is able to produce a visualization figure that is replicate the actual people's work mobility for the year 2022. After evaluating each model separately, it turned out that the best result was for the Neural Network model with three hidden layers, with 200, 20, and 200 neurons, respectively. For the activation function, ReLu was used with Adam as an optimizer, and the maximum number of iterations was 30.

Based on that, a time series data visualization was produced for same five months in different years. Three figures are for the actual workplace mobility percentage changes for 2020, 2021, and 2022, and the fourth figure, in the right corner below, was produced based on Neural Network approach that predicted 2022 workplace mobility changes based on the data of the two previous years, 2021 and 2020. In the figures, the y-axis is mobility changes, and the x-axis is the corresponding months. The Python visualization library used is Matplotlib. To increase figures readability, each value of mobility change was flipped. For instance, a decrease in mobility of -30% was multiplied by -1 to become 30%. This made the figures reads bottom-up. Moreover, different contrasted colors utilized with different layered hues. This eases the readiness and makes the figure more appealing. Each subplot was given a title and legend based on the information presented independently while a global title is also included.

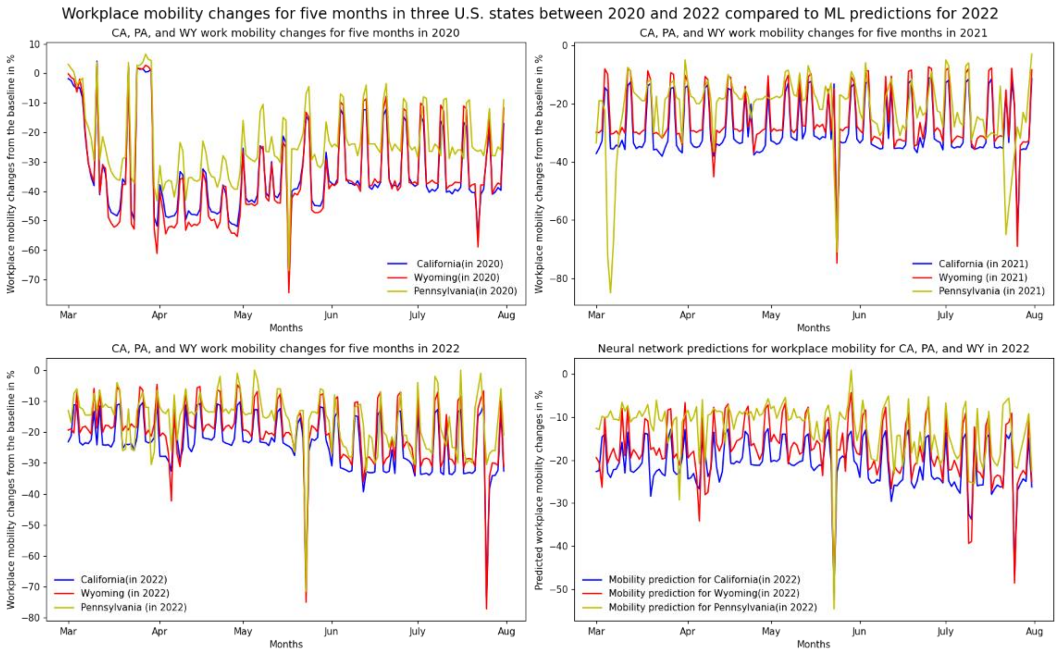
**The results show the following:**

- The ML model (right bottom figure) was able to predict changes for the year 2022, with an accuracy (R2) of 63.2%.
- Visualizing and comparing results was helpful in identifying the overall trend in the three years.
- In the years 2021, Wyoming had the largest decline in mobility, around 80%.
- In 2020 and 2022, Pennsylvania had the greatest decline, which was around the beginning of May.
- The mean of work mobility changes indicates those states are still recovering towards normal, which is 0% on the figures.
- It seems people still work in a hybrid mode as of the date above.
- COVID-19 in 2021 and 2022 seems to be not only the factor affecting work mobility.
- Data shows that over the covered period and in all three states, mobility dropped around the end of May. In all three states. This might have other contributing factors other than COVID-19 such summer break.
- Based the average line (the horizontal line), each year the average decreases towards zero, which means we are progressing the normal life again.

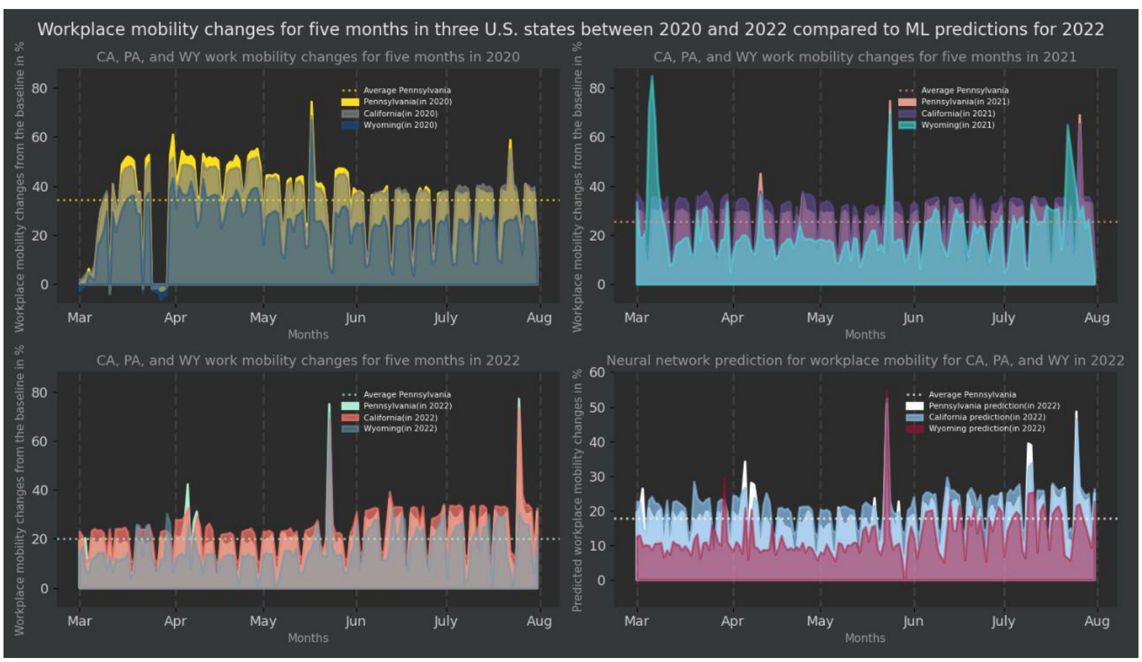
**Comparison - first vs. final version**

In this first version, the figure represents the data in both negative and positive data points. This is not the case in the final version, since most of the values are negative, it was decided data points need to be multiplied by -1. As a result, this made the figure more readable. Also, the background of both each subplot and the figure were colored. To make it charming further, coloring and layering schemes were implemented. Since our main interest is in Pennsylvania, it was given the solid and brightest color and placed in the background. The second layer was California, the largest state, and was made transparent and positioned as a middle layer whereas, for Wyoming, which is the smallest state, the color was also transparent and place in the foreground. This made the figures more attractive and easy to read. Moreover, the first version did not include the average line (the horizontal line in the middle of each figure) which was calculated based on the Pennsylvania average for that year. For that we can see over the course of the three years the average started to decline towards zero, which indicates a normal life recovery after COVID-19.

- ***The first version of the work mobility changes figure:***




- ***The final version of work mobility changes figure after enhancement:***




**GitHub link (code, data, and documentation):**

- https://github.com/abdulrahmanhi/Information_Visualization_Final_Project

**Reference**

[1] Google LLC *"Google COVID-19 Community Mobility Reports"*.
https://www.google.com/covid19/mobility/ Accessed: (Sep. 15, 2022).