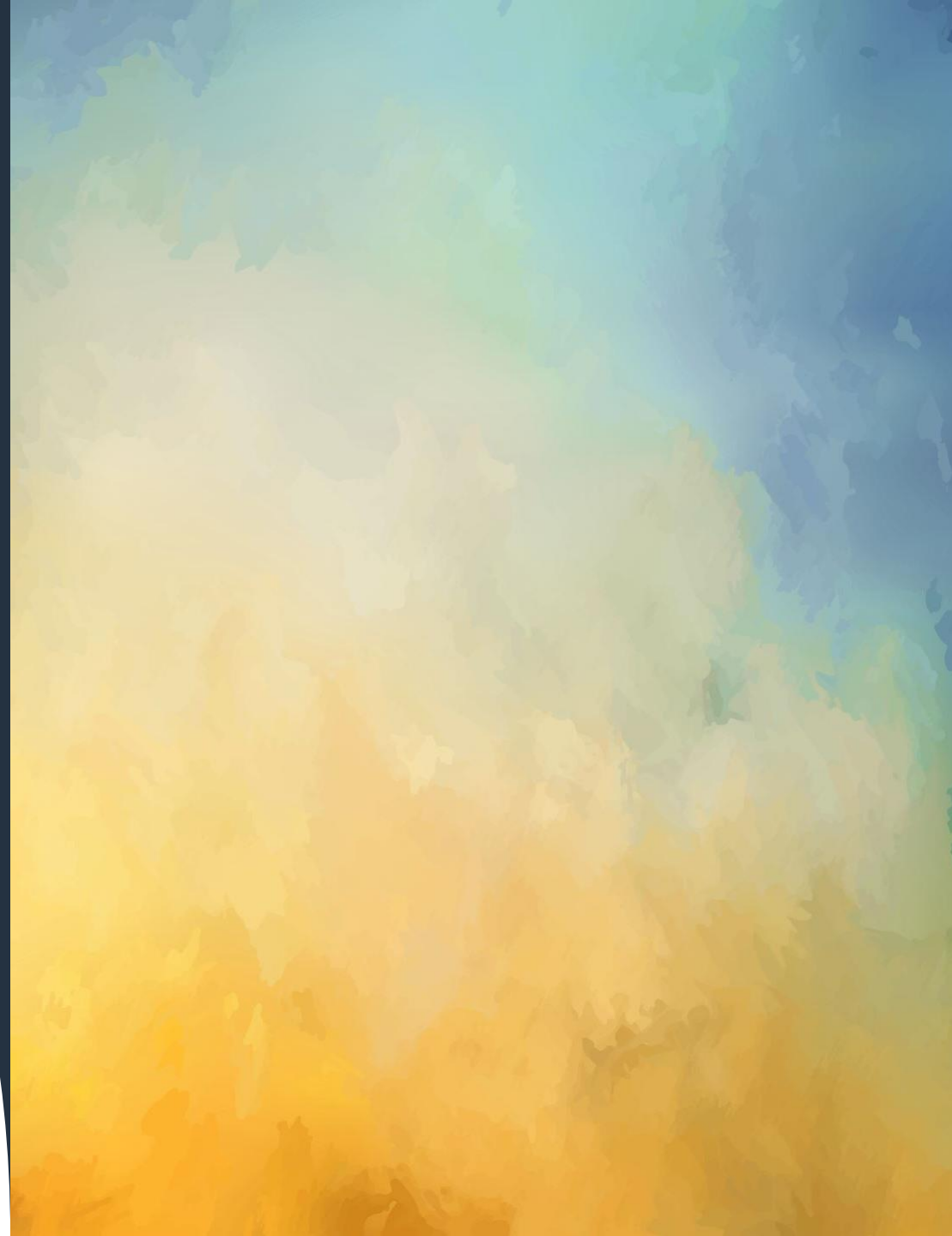
Guidelines for Digital skills Lab

Prof.ssa Cristina Pronello

Deepan Anbarasan



Exercises

1) Preliminary analysis
2) The where and when of the mobility - OD matrices
3) Relation between Public Transport lines and Origins/destinations of trips – Is there a relation?
4) Parking duration and business model

Data setup

- Download micro-mobility data for Austin, Texas from here or access the data using API from here.
- Download Public transport information for Austin from here.
- Download micro-mobility data from Chicago from here or access the data using API from here.
- Download public transport information for Chicago from here.
- Download administrative boundaries information for Chicago from here.
- Exercises 1 and 2 will be focused on the Austin dataset. Exercise 3 and Exercise 4 will involve the Chicago dataset.
- The programming language of choice for the exercises is Python.
- The choice of how to store the data for future access is left to the students: csv files, MongoDB, PostgreSQL, MySQL etc.,
- Suggestions :
  - If csv files are used as is – to avoid Out-of-memory errors – use the pandas library's read_csv function and use the *chunksize* parameter to read the csv file in chunks or nrows parameter to read a specific number of rows.
  - For checking if the read_csv has worked. df.head() command can be used.

Exercise 1 – Preliminary analysis

Before answering any questions, some initial data cleaning might be required - Removing empty rows or rows missing essential data.

Answer the following questions based on the data setup and using python:

- The number of records before and after data cleaning. What were the types of "bad data" that needed to be cleaned ?
- When did collection of data start(start-date and time) for micro-mobility dataset and what is the most recent date and time available.
- Number of records per year and month in micro-mobility dataset – Use bar plots to display patterns using the plot() function of pandas.
- Number of unique vehicles by year and vehicle type – identify patterns e.g.,:
  - Are there more vehicles as years go on ?
  - Is there some change in usage patterns among different days of the week , months is there a trend – seasonal or weekly ?
  - Are there any trends based on the gender and age of the user ?
- Calculate the number of trips for each vehicle in the database and produce a histogram for it.

Exercise 2 – OD matrix

- Compute then the O-D matrix, i.e., the number of bookings starting in council district *i* and ending in council district *j*. Try to visualize the results in a meaningful way. If the students need a challenge, they can use census tracts instead of council districts.
- Prepare OD matrices for different hours of the day and different age groups(3 OD matrices for each age-group for peak hours traffic and 3 for non-peak hour traffic). Are there any periodicity or trends noticed? Is there a difference between the OD matrices for different age groups?
- Based on your observation, visualise selected OD matrices that show some trends/periodicity on a map.
- If a challenge is needed, students can also attempt to create a flowmap for the OD matrices.

Exercise 3 – Relation to Public transport line

- Since we don't have detailed location data for bookings in Austin, Chicago data can be used for this exercise. Filter the data to consider only one month of data – as the data analysis for this exercise might take a long time depending on how it's implemented.
- Import the public transport lines(as kml files) as one layer on QGIS or ArcGIS, and the Micro-mobility data in csv files as another layer.
- Visualising the two layers, record your observations about any patterns on interaction between micro-mobility and public transport – complementary, competing etc.,.

Exercise 4 – Parking duration and costs

- What is the average parking duration of vehicles in different council districts or wards? – Display this info using a map.
- Do an overlapping visualisation with OD matrices and average parking durations across the different parts of the city for different hours of the day.
- Calculate the revenues (different streams of revenue for a micro-mobility company) and try to calculate the costs (both capital and operational) to understand the business model – based on the guidance provided by Prof.Pronello.
- Chicago data is more granular and better for this exercise but if you have trouble using Chicago data due to processing power constraints – use Austin data.
- If you use Chicago dataset:
  - Associate each trip in the dataset to an origin and destination ward by combining the trip information with the information about the wards. Check which wards the FROM_LOCATION and TO_LOCATION fields belong to.
  - Compute then the O-D matrix, i.e., the number of bookings starting in ward *i* and ending in ward *j*. Try to visualize the results in a meaningful way.