**Homework Week 2**

This week’s homework will be purely Project based. You need to work as a group and the homework will be submitted by one of the members of your group. List the other members of your group in the document, so that your instructor can mark every student.

**Group members:(Group 6)**

1. Dinara Akmurzina
2. Marina Zviagina
3. Magdalena Krupa
4. Sharyn McPherson
5. Vanessa Chan
6. lamiaa Bahgat

**Assigned instructor:** Andreea

**Question 1 [20 points]**

What is your project question(s) and the problem it tackles?

* You will be marked on how realistic the problem is given the project timeline, as well as the fit with the subjects covered throughout the Data Specialisation and your personal, common interests in the topic.

Project Question: Does the presence of bike stations and docks in Austin help promote bike usage that enables cleaner transportation methods?

Problem: The problem tackled by this project is to help determine if bike stations, docks and the bike usage in Austin contribute to promoting cleaner transportation methods. The problem will be assessed by analysing various factors, these include:

* identifying the most popular routes taken by commuters.
* bookings taken in each station.
* ratio of visitors and commuters
* subscription types and optimisation
* duration of trips
* travel behaviours
* activity per bike station
* bike usage - day,week and month.
* peak/off-peak usage

The above factors will help us explore and analyse different trends and patterns throughout Austin and show if the utilisation of bike stations is a popular choice of transportation within this location. To address the problem the project will involve datasets that include information on bike stations and trips within Austin. The dataset that will help determine analysis for our project question should contain data on bike station locations, the number of docks available, trip details including start times, trip duration and other user information. Statistical methods and visualisations will also be explored to tackle the problem our project presents.

By investigating these factors and understanding bike usage patterns, this project aims to provide insights into the effectiveness of Austin's bike-sharing infrastructure in promoting cleaner transportation alternatives. This information can be useful for city planners and policymakers to make informed decisions regarding the expansion of bike-sharing programs and the development of sustainable urban transportation systems.

*·*

**Question 2 [20 points]**

Explain your target audience. Who could be interested in reading your final report and for whom will your project be useful? Assess the level of expertise in relation to data science of your audience (for example, how technical should your report/presentation be?).

* You will be marked on correctly identifying your audience and the technical level of your presentation and report.

Our target audience for this project includes:

1. ***Environmental and sustainability organisations*:** groups that are focused on reducing carbon emissions could find our report interesting to help promote any positive environmental agendas, for example sustainable transportation options or accessibility of cleaner transportation methods within Austin. These organisations could also use our findings to promote benefits of cleaner transportation initiatives in other locations,cities and countries. Examples include local non-profit organisations ([Most Popular Austin, TX Nonprofits and Charities | Donate, Volunteer, Review | GreatNonprofits](https://greatnonprofits.org/city/austin/TX/category:environment/sort:review_count/direction:desc)) in Austin and international organisations such as Sustrans.org.uk.
2. ***Bike advocacy groups*:** Organisations or individuals dedicated to promoting bike usage as a mode of transportation, such as [Bike Austin](https://www.bikeaustin.org/) could be interested in our findings. These groups promote active transportation and encourage bike usage, aiming to create a more extensive and safer bicycle infrastructure in their communities.
3. ***Transportation authorities and officials*:** Organisations that are responsible for more efficient transportation systems can use our findings to explore patterns and adopt more efficient transport options and accessibility. Examples include the Sustainability and Transportation departments at AustinTexas.gov

[Sustainability | AustinTexas.gov](https://www.austintexas.gov/department/sustainability),

[Transportation | AustinTexas.gov](https://www.austintexas.gov/department/transportation)

[Transportation | AustinTexas.gov](https://www.austintexas.gov/department/transportation) .

1. ***Residents, community members and the general public:*** Anyone interested in living in Austin or planning to visit there may find our report useful. Our findings could provide information and insights into cleaner transportation methods for everyday use or for their next trip to Austin.
2. ***Travel agencies and tourists***: People interested in planning travels through the city, including popular tourist areas, can benefit from our report to explore bike-friendly routes and locations.
3. ***Students, researchers, and academics***: Scholars, researchers, and students studying urban planning, transportation, and sustainability might be interested in exploring the impact of various factors, such as bike stations and docks, on transportation patterns and sustainability goals within the Austin community.

The report should have a moderate technical level: it should possess efficient understanding of analysis concepts, techniques and data visualisation but with simplicity for our audience to be able to understand and view with ease. Having a solid technical foundation that enables a strong data analysis for our findings will have a greater impact on our audience. Having both written and visual communications will help our report be more accessible to a broader audience.

**Question 3 [30 points]**

What data sources will you need to answer your project question(s)? Describe any potential issues you can have with the datasets and how will you overcome this:

* For example, will the data you find only cover particular geographical areas? Will you need to combine multiple datasets to overcome this?

***Data sources:***

1. Dataset 1:Bike Trip data - this dataset contains information on the bike trips taken from December 2013 to March 2013.
2. Dataset 2:Bike Station data - this dataset contains information on all active and inactive bike stations.

***Potential issues:***

1. *Incomplete or inconsistent datasets*

Our strategy to overcome this problem is by assessing the data and either fill in missing values if appropriate, or drop rows/columns if necessary.

1. *Inaccurate data*  
   For example, if a bike station address is not correct, or trip duration recorded incorrectly, by cross-examining other data provided in the dataset we can then amend the inaccurate data. If not, dropping rows or columns if necessary is another solution to this issue.
2. *Lack of information in data.*

For example, absence of information provided on when the bike stations were closed, or when new bike stations were opened. May need to conduct extra research to fill in the gaps to supplement the dataset’s information.

1. *Data privacy.*

The dataset might include sensitive information about users, such as names, email addresses, or other personally identifiable information (PII). It is essential to handle such data carefully to protect users' privacy and comply with data protection regulations. To overcome this, anonymize any sensitive data before analysis.

1. *Data biases*

The data may have biases that could affect the generalisability of the analysis, such as overrepresentation or underrepresentation of certain user groups or geographic areas. To mitigate this, we should explore the dataset to identify any potential biases and discuss these limitations in our analysis.

1. *External factors*

The dataset might not capture external factors that could affect bike usage, such as special events, or changes in transportation infrastructure. This could limit our ability to draw accurate conclusions about the impact of bike stations and docks on promoting sustainable transportation.

1. *Changes in data collection methods.*

If the data collection methods have changed over time, it could impact the consistency and comparability of the data. To address this, we will investigate the documentation of the dataset to understand any changes in data collection methods and consider adjusting our analysis to account for these changes.

**Question 4 [30 points]**

Describe the team approach to the project work:

* how are you planning to distribute the workload and how are you planning to work on your project;

**Distribution of workload:**

*1.Plan regular (weekly) team meeting via zoom/meet*

Enables the team to define goals, objectives and deliverables before analysis and reporting. Discussion of specific activities and task needed for project completion - address the research question and adjust to our projected timeline. Being consistent with each others work, keeping a diary of meeting dates. Keeping track of meeting minutes taken to summarise progress and points raised during discussions.

*2.Communicate via slack:*

Regular communication and check-in with the team via our group slack chat. Daily updates, any challenges or questions communicated to maintain continuous collaboration.

3.*Task Allocation*:

Use Trello to manage workflow of the project and easily get an overview of the progress of the project. Use Trello for task tracking and allocation - our team can track the progress of all tasks as well as add and update tasks. Allocations are based on strengths, skills and particular interest in the task. Check and review each other's work on GitHub and on our shared Google Colab Notebook.

* what are your team’s strengths and weaknesses;

**Strengths:**

1. *Fresh perspective*

Having a team from different backgrounds and perspectives will bring original ideas and fresh approaches to the project.

1. *Willingness to learn*

Our team is eager to learn and expand their knowledge to be able to complete the project to a high standard.

1. *Proactive*

Our team is taking a proactive approach to all tasks needed to progress forward efficiently

1. *Attention to detail*

Members of our team that are detail-oriented can utilise this skill to enhance our analysis and catch potential errors.

1. *Communication skills*

Articulation of ideas, communication of feelings and actively listening to other team members to push forward within tasks.

1. *Well organised*

We are able to manage our time, tasks and timeline effectively making smooth progress and efficient completion of the project.

1. *Open minded approach*

Our approach allows for team members to be considerate to different perspectives, to change and challenge any ideas or suggestions presented.

**Weaknesses:**

1. *Limited skills*

Our team has a varying level of expertise that could impact the effectiveness of the project. Collaboration, sharing of knowledge and external research would help overcome this weakness.

1. *Time constraints*

Our team members have other commitments - including CFG coursework work obligations and other personal responsibilities. Limiting the amount of time dedicated to the project, prioritising other commitment and more efficient time management would help overcome this weakness.

1. *Limited availability* - everyone has different schedules making collaboration more challenging. Using open communication, slack and trello has helped overcome this weakness.
2. *Limited resources* - large datasets are causing our computers to run slower. Using APIs or other tools to help overcome resource limitations.

* how are you managing your code;

1. *Upload individual codes to Github - as a pull request*

Each team member feels confident to work on their assigned task and upload their code to a shared repository on GitHub.

2. *Select code to merge into branch*

During team meetings, we can discuss and evaluate the code submitted.

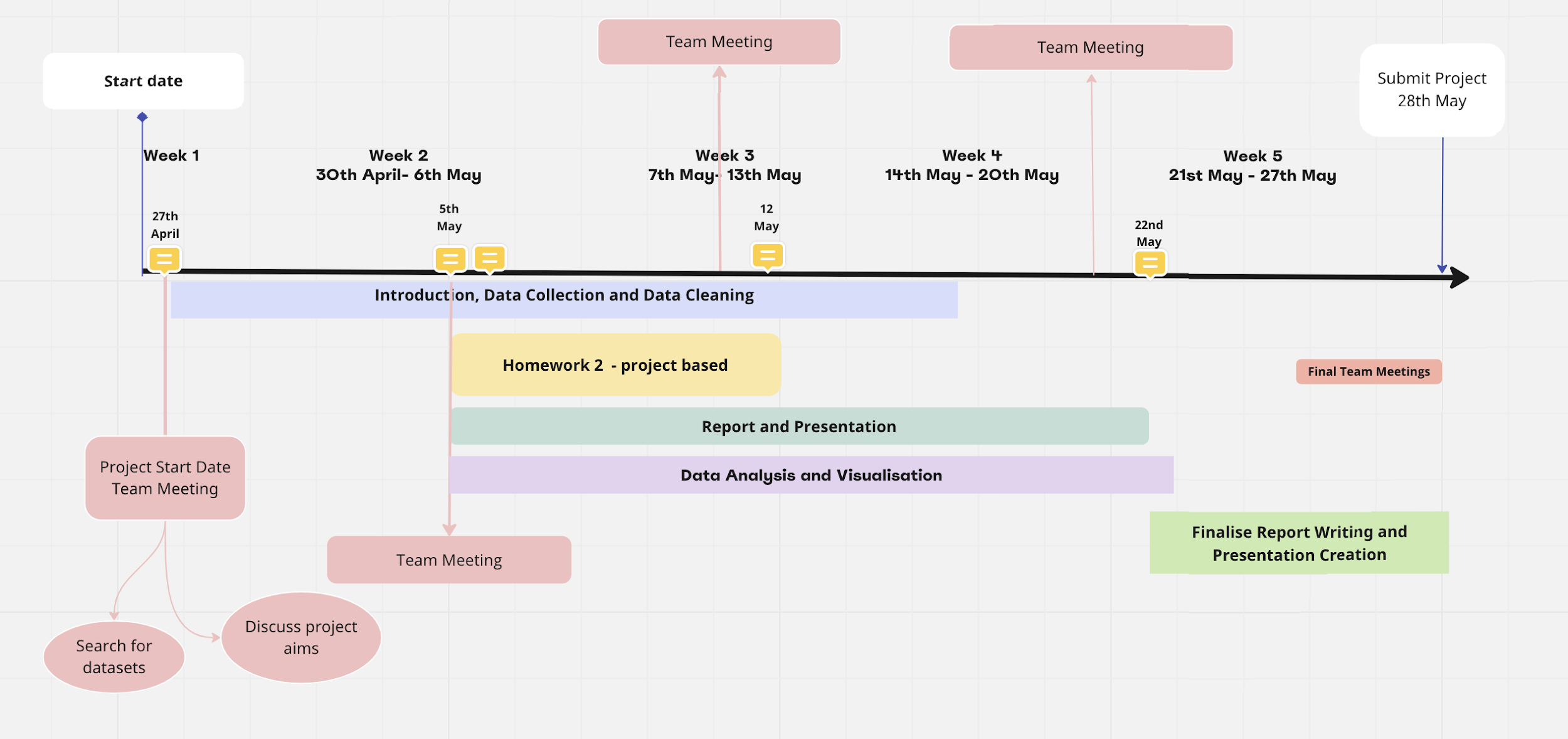
3. *Combine and discuss codes in meetings*

During our team meetings and communication through slack we can discuss the progress and any updates related to the code.

4. *Shared folder for project essential files*

To maintain organisation and easy access of our project files, we have organised a shared folder. Our datasets and documents are stored here.

* include an expected timeline of the project



Week 1

27th April - Project Start Date

1. Initial team meeting - discuss project subject, questions, aims and objectives
2. Search and select datasets

Week 2

30th April - 6th May

1. Start on introduction to project
2. Work on aims and objectives
3. Work on roadmap
4. Data collection
5. Data cleaning
6. 5th May - Start project based homework
7. Collaboration team meeting

Week 3

7th May - 13th May

1. Continue on with data collection
2. Continue on with data cleaning
3. Complete and submit project based homework
4. Work on data analysis and visualisation
5. Collaboration team meeting

Week 4

14th May- 20th May

1. Work on data analysis and visualisation
2. Work on project report and presentation
3. Collaboration team meeting

Week 5

21st May - 27th May

1. 22nd May - Complete data analysis and visualisation
2. Continue work on report and presentation
3. 22nd-27th finalise report writing and presentation
4. Final team meeting(s)

28TH MAY - SUBMISSION