Assignment 1

## Tisha Radia

#### a1952871

**1. Initial Big Question and Societal Relevance**

The foundation of contemporary existence is urban mobility, which makes possible the social, educational, and economic interconnectedness that cities depend on. In order to service a diverse population with a range of needs, public transportation systems in particular need to run effectively and dependably. However, a variety of external factors interact to impact the use of public transportation; weather and public events are two significant but little-studied variables.

This project seeks to answer the following **big question**:

**“How do combinations of weather conditions and public events affect public transport usage in cities?”**

This question is of **broader societal relevance** because:

* **Optimising Service Delivery:** Understanding usage trends under various weather and event conditions allows transportation authorities to dynamically modify frequency, staffing, and capacity, decreasing congestion and enhancing user experience.
* **Infrastructure Planning and Investment:** The insights gained can be used to influence strategic infrastructure decisions, such as building shelters at bus stations that are regularly utilized during inclement weather or enhancing lighting and safety around venues with heavy evening traffic.
* **Urban Resilience and Sustainability:** Given the growing unpredictability of the climate, it is essential to comprehend how bad weather impacts the use of public transportation. Even in difficult circumstances, this can promote modal changes away from the use of private vehicles and boost climate adaption initiatives.
* **Policy and Decision-Making:** To efficiently distribute resources and prepare for periods of high consumption, governments and urban planners need data-driven projections. This includes taking into consideration the consequences of sporting events, concerts, public holidays, and heat waves.
* **Equity and Accessibility:** Regardless of inclement weather or disruptions from events, some communities may depend more on public transportation. Cities may guarantee transportation fairness for vulnerable groups by modelling these factors.

**2. Big Data Sources to Be Collated**

A variety of datasets from other disciplines will be combined to provide an answer to the question. These datasets demonstrate volume, diversity, velocity, and veracity issues, which are characteristics of Big Data as defined in Module 1.

**Primary Data Sources**

1. **Adelaide Metro Public Transport Statistics**
   * Source: <https://adelaide-metro-public-transport-statistics-unisthaus.hub.arcgis.com/>
   * Type: website
   * Description: Includes boardings by mode (bus, train, tram), time, stop, and region across multiple years.
2. **Weather Data from Bureau of Meteorology (BoM)**
   * Source: <http://www.bom.gov.au/climate/data/>
   * Type: Time-series
   * Description: Daily and hourly weather data (temperature, rainfall, wind, etc.) by location and time.
3. **Events Calendar (South Australia)**
   * Source: Local council websites, Eventbrite API, or tourism databases
   * Type: Categorical/tabular
   * Description: Events by date, type, and location (e.g., concerts, sports, festivals).
4. **School and Public Holiday Calendar**
   * Source: SA Government holiday schedules
   * Type: Structured
   * Description: Key non-working days that may influence travel patterns.

**3. Strategy for Data Processing and Integration**To ensure uniformity and analytical robustness, the data collected from multiple sources will undergo a methodical pre-processing and cleaning phase, primarily using R (dplyr, lubridate, tidyr) or Python (Pandas, NumPy, scikit-learn). Because of the complexity and heterogeneity of the datasets, the cleaning process will incorporate the following essential steps: event calendars, weather records, and transport logs:

* **Datetime Alignment:** Depending on each dataset's availability and resolution, all data sources will be synchronized on a common time axis, preferably at the daily or hourly level. For instance, to guarantee time-based consistency across sources, hourly weather measurements and boarding data will be synchronized. Capturing quick behavioral changes brought on by weather variations or event start/end times requires temporal granularity.
* **Managing Missing Values:** Context-aware imputation techniques will be used to handle missing data points, such as missing weather observations, gaps in ridership logs, or incomplete event listings. Missing values, where suitable, could include:
  + For time-series weather data, interpolation or rolling means are used to impute
  + For ridership counts, carried forward/backward (LOCF/FOCF), or
  + If the absence substantially skews the analysis, it is flagged for exclusion.
* **Geospatial Mapping and Location Granularity:** Using postcode, latitude/longitude information, or shapefiles for spatial joins, weather stations and event locations will be mapped to the closest transit stops or suburbs. This phase makes it possible to combine datasets that overlap spatially but do not have a common geographic identification. It guarantees precise mapping of weather-sensitive or event-driven behaviors to particular transportation network segments.
* **Outlier Detection and Annotation:** Unusual trends will be identified as contextual outliers, such as sharp declines during transport strikes or sharp increases during significant occasions or holidays. Instead of being eliminated, these will be marked so that they can be taken into consideration as significant anomalies or taken into consideration when making model improvements in later phases of the research.
* **Data Integration:** Composite keys, such as the following, will be used to integrate the datasets:
  + Date or time (to align weather, ridership, and events),
  + location indicators, such as stop ID, postcode, or suburb, and
  + Where appropriate, the mode of transportation (bus, rail, tram).

The combined dataset will create a multi-index time series that is organized to facilitate predictive modeling, visualization, and exploratory research.  
  
The basis for finding correlations, visual patterns, and maybe training algorithms to predict public transportation usage under various weather and event-driven situations will be this pre-processed and unified information.

**4. Data Limitations and Refinement**

**Potential Data Limitations**

* Event data may be incomplete or vary in detail across regions.
* Some stops may not have consistent boarding records.
* Weather stations may not exactly match event/stop locations (requiring geographic interpolation).

**Refined Question (if needed)**

If event data is sparse or unstructured, the question may be narrowed to:

“How does adverse weather affect transport usage patterns across different suburbs of Adelaide?”

**5. Backup Question and Data Source**

**Backup Question**:

“What patterns in public transport usage can be explained by weather and holiday schedules alone?”

**Backup Data Sources**:

* Bureau of Meteorology (weather)
* Adelaide Metro public transport usage
* SA school/public holidays

This backup plan excludes event data and simplifies the modelling to weather and date-based analysis only.

**6. Next Steps**

* Complete the compilation of data from the four sources.
* Datasets for the two-year sample period (2022–2024) should be cleaned and integrated.
* Execute exploratory visualizations, such as correlation heatmaps and line charts.
* Create a baseline model, such as time-series forecasting or regression.
* Look for trends, patterns, and abnormalities to generate more hypotheses.

**7. References**

* Adelaide Metro Public Transport Statistics Hub (n.d.) *Adelaide Metro Public Transport Statistics*. [online] Available at: <https://adelaide-metro-public-transport-statistics-unisthaus.hub.arcgis.com/> [Accessed 12 June 2025].
* Bureau of Meteorology (n.d.) *Climate Data Online*. [online] Available at: <http://www.bom.gov.au/climate/data/> [Accessed 12 June 2025].
* Government of South Australia (2025) *Newsroom – Premier of South Australia*. [online] Available at: <https://www.premier.sa.gov.au/media-releases> [Accessed 12 June 2025].
* 7NEWS Adelaide (@7newsadelaide) (2025) *[Instagram profile]*. [Instagram] Available at: <https://www.instagram.com/7newsadelaide/> [Accessed 12 June 2025].