Data Management Plan

Content

Table of Contents

1. Datasets Overview

1.1 Iteration 1

Name	Physical Acess	Frequency of Source Updates	Frequency of System Updates	Granularity	Copyright / Licensing Details
Google Maps API	API	Real-time	As needed	High	Google Maps APIs Terms of Service
Sport and Recreational Facilities List	CSV	Annually	As needed	Medium	Creative Commons Attribution 4.0 International
Crime Incident Dataset	CSV	Annually	As needed	High	Creative Commons Attribution 4.0 International
OpenWeatherMap API	API	Real-time	As needed	Medium	OpenWeather APIs Terms of Service
Public Playground Locations	CSV	Annually	As needed	Medium	Creative Commons Attribution 4.0 International
Landmark Dataset	CSV	Annually	As needed	Low	Creative Commons Attribution 4.0 International
Victoria Road Crash Data	CSV	Annually	Quarterly	High	Creative Commons Attribution 4.0 International

1.2 Iteration 2

1.2.1 Open Data

Name	Physical Access	Frequency of Source Updates	Frequency of System Updates	Granularity	Copyright / Licensing Details
Australian Government's Physical Activity Guidelines for Children and Young People	Government website	Annually or when updated	As needed	Medium	Creative Commons Attribution 4.0 International
World Health Organization - Physical Activity Guidelines	WHO website (downloadable PDFs, online access)	As updated by WHO (irregular)	As needed	Medium	Creative Commons Attribution 4.0 International
Australian Government's Physical Activity Guidelines and AAP	Websites: Health.gov.au, AAP	Annually or when updated	As needed	Medium	Creative Commons Attribution 4.0 International
U.S. Department of Health and Human Services - Physical Activity Guidelines	HHS website	Every 10 years	As needed	Medium	Public domain (US government)

1.2.2 Team-Generated Data

Name	Physical Access	Frequency of Source Updates	Frequency of System Updates	Granularity	Copyright / Licensing Details
Team-uploaded photos for scavenger hunt game.	Photos uploaded via the web app	As needed	As needed	High	Internal use, no storage of personal data

1.3 Iteration 3

1.3.1 Open Data

Name	Physical Access	Frequency of Source Updates	Frequency of System Updates	Granularity	Copyright / Licensing Details
First Aid and Safety Guidelines for Children	Government website (PDF)	Annually or as needed	As needed	Medium	Creative Commons Attribution 4.0 International
Australian Government's Physical Activity Guidelines for Children and Young People	Government website	Annually or when updated	As needed	Medium	Creative Commons Attribution 4.0 International
Australian Red Cross First Aid for Children	Downloadable from the Red Cross website	Annually or when updated	As needed	High	Public domain

1.3.2 Team-Generated Data

Name	Physical Access	Frequency of Source Updates	Frequency of System Updates	Granularity	Copyright / Licensing Details
Introduction to website for Chat Bot Training	Internal repository	As required by team	As needed	Medium	Internal use

2. Data Usage

2.1 Iteration 1

- **Sport and Recreational Facilities List**:
  This dataset contains information on various sports and recreational facilities in Melbourne, including their location and types of sports available. It will be used to map out safe and accessible physical activity locations for immigrant children. The dataset will allow immigrant children to find nearby facilities that meet their specific needs and preferences.
- **Crime Incident Dataset**:
  The dataset provides annual crime statistics across various postcodes and suburbs in Melbourne. It will be utilized to assess the safety of different areas by analyzing historical crime data. This information will be visualized on the map to guide immigrant children away from higher-risk areas, enhancing their safety during physical activities.
- **Playground Dataset**:
  This dataset includes locations and names of playgrounds in Melbourne. It will be integrated into the map to help immigrant children find nearby playgrounds suitable for children's physical activities. The dataset will also contribute to identifying safe areas by cross-referencing with crime data.
- **Landmark Dataset**:
  This dataset comprises various landmarks across Melbourne. It will be used for orienting immigrant children and providing contextual information on their surroundings. While landmarks are not directly related to safety, their inclusion can help immigrant children better navigate and choose familiar or well-known locations for activities.
- **Victoria Road Crash Dataset**:
  This dataset contains detailed information about road crashes in Melbourne, including severity and locations. It will be used to highlight areas with high accident rates, helping immigrant children avoid potentially dangerous routes. This information will be displayed on the interactive map to further enhance the safety aspect of the tool.
- **Google Maps API**:
  The API will be used to fetch the latitude and longitude of immigrant children' current locations as well as the destination locations. It will be essential for creating interactive route maps that guide immigrant children to their chosen physical activity locations.
- **Real-time Weather Forecast API**:
  The weather API will provide real-time weather conditions, including precipitation, temperature, and UV index. This data will be crucial in helping immigrant children decide the best time to engage in outdoor activities, ensuring they are prepared for the weather conditions they might encounter.

2.2 Iteration 2

- **Team-uploaded photos for scavenger hunt game**:
  This dataset consists of photos taken by the development team from various parks and landmarks in Melbourne. These photos will be used in the backend to compare and verify the locations of user-uploaded photos during the scavenger hunt. By analyzing and matching user-uploaded images with the reference photos taken by the team, the system can confirm if the user has successfully found the correct location. This ensures accuracy in the scavenger hunt and enhances the gameplay experience while keeping the user's data private.
- **Australian Government's Physical Activity Guidelines for Children and Young People**:
  This dataset includes official guidelines on the recommended levels of physical activity for children and young people, based on age and development stages. It will be used to inform parents about whether their children are meeting the necessary physical activity requirements. The dataset will provide structured recommendations based on survey responses, helping parents ensure their children engage in adequate physical activities.
- **World Health Organization - Physical Activity Guidelines**:
  This dataset provides global guidelines on physical activity for children and adolescents. It will be used to compare and complement Australian guidelines, offering parents a broader perspective on healthy activity levels for their children. The guidelines will be incorporated into personalized recommendations for both outdoor play and structured activities.
- **Australian Government's Physical Activity Guidelines (Health.gov.au) and American Academy of Paediatrics (AAP)**:
  This dataset combines recommendations from both the Australian Government and the AAP regarding physical activity and screen time for children. It will be used to educate parents on how to balance their child's screen time with outdoor activities. The recommendations will help parents monitor and reduce sedentary behaviour, encouraging children to meet their daily physical activity targets.
- **U.S. Department of Health and Human Services - Physical Activity Guidelines**:
  This dataset provides long-term physical activity guidelines aimed at improving the overall health of children. It will be used to provide strategic recommendations for structured physical activities, such as sports or school-based physical education programs. The dataset will serve as a reference for developing tools that help parents track and promote regular physical activity in their children.

2.3 Iteration 3

- **First Aid and Safety Guidelines for Children (Health.gov.au)**:
  This dataset will provide structured first aid guidelines and safety practices that will be incorporated into the educational content of the app. The content will focus on basic first aid and injury prevention techniques, helping immigrant children remain safe while engaging in physical activities. These guidelines will be presented in a child-friendly format, empowering children with practical skills to handle minor accidents and injuries.
- **Australian Government's Physical Activity Guidelines for Children and Young People**:
  This dataset provides authoritative guidelines on the recommended levels of physical

activity for children based on their age. It will be used to create personalized activity recommendations, and the dataset will be integrated into the AI-powered chatbot to ensure that it provides accurate and age-appropriate physical activity guidance for children.

- **Australian Red Cross First Aid for Children**:
  The Australian Red Cross dataset offers valuable insights into first aid techniques that are specifically tailored for children. This data will be integrated into the app's educational modules to provide a more visual and interactive approach to learning about first aid. Immigrant children will be able to access simple and clear instructions for dealing with common injuries like cuts, bruises, and burns while playing in parks.
- **Introduction to website for Chat Bot Training**:
  This dataset contains internally developed documentation about the features and structure of the website. It will be used to train the AI-powered chatbot, ensuring that the chatbot can effectively answer questions related to the platform's navigation and features. The data will help provide clear, accessible explanations for children who need assistance using the website, enhancing user experience and engagement.

3. Data Preparation

3.1 Cleaning and Wrangling

Each dataset underwent a rigorous cleaning process. We removed unnecessary columns, standardised formats, and ensured that only the most relevant data was retained for analysis. And here is some special step below:

1. **Accident Dataset:**

```
# Sort by total_severity in descending order
sorted_severity_per_postcode = severity_per_postcode.sort_values(by='total_severity', ascending=False)

print(sorted_severity_per_postcode.head(20))

```

```
     postcode  total_accidents  total_severity
151      3175             3227            8311
0        3000             2212            6021
677      3977             2322            5989
282      3350             1909            5250
22       3029             1943            5108
23       3030             1751            4587
55       3064             1676            4445
99       3121             1561            4180
62       3072             1452            3928
13       3020             1492            3909
```

We merged the accident and node location datasets using ACCIDENT_NO as the key, retaining essential columns like SEVERITY, LATITUDE, LONGITUDE, and POSTCODE_CRASH. After merging, we checked for and removed 2,718 duplicate entries to ensure data accuracy.

Next, we aggregated the data by postcode, calculating the total number of accidents and their severity for each area. The aggregated data was then sorted by total_severity in descending order, highlighting the top postcodes with the highest severity levels. This process helps identify areas with significant safety concerns for further analysis.

2. **Facility Dataset**

```python
# Replace NaN values with 'Unknown' in the 'sports_played' column
facility_data['sports_played'] = facility_data['sports_played'].fillna('Unknown_event')

# Verify the changes
print(facility_data.head())

```

```
  facility_id       facility_name  suburb   pcode   latitude   longitude  \
0  MITCHE3674  Vermont South Club  BRIGHT  3741.0 -37.851683  145.180228
1  MITCHE3674  Vermont South Club  BRIGHT  3741.0 -37.851683  145.180228
2  MITCHE3674  Vermont South Club  BRIGHT  3741.0 -37.851683  145.180228
3  MITCHE3674  Vermont South Club  BRIGHT  3741.0 -37.851683  145.180228
4  MITCHE3674  Vermont South Club  BRIGHT  3741.0 -37.851683  145.180228

              sports_played
0             Unknown_event
1  Australian Rules Football
2                   Cricket
3                   Netball
4          Tennis (Outdoor)
```

```python
# group the sports located ar the same facility
facility_data_grouped = facility_data.groupby(
    ['facility_id', 'facility_name', 'suburb', 'pcode', 'latitude', 'longitude']
).agg({
    'sports_played': lambda x: ', '.join(x.dropna().unique())
}).reset_index()

# Verify the changes
print(facility_data_grouped.head())

```

```
   facility_id                    facility_name  suburb   pcode   latitude  \
0  ALPINE10916   Bright Skate Park (lions Park)  BRIGHT  3741.0 -36.728932
1  ALPINE10917                Bright Bowls Club  BRIGHT  3741.0 -36.731007
2  ALPINE10918              Bright p-12 College  BRIGHT  3741.0 -36.733701
3  ALPINE10919  Bright Community Sports Stadium  BRIGHT  3741.0 -36.733294
4  ALPINE10920             Bright Sports Centre  BRIGHT  3741.0 -36.728695

    longitude                                       sports_played
0  146.970354                                             Skating
1  146.969729                                          Lawn Bowls
2  146.962508                                  Cricket, Soccer
3  146.963623                                          Basketball
4  146.958053  Fitness / Gymnasium Workouts, Squash / Racquet...
```

In this data preparation step, we first addressed missing values by replacing NaN entries in the `sports_played` column with `'Unknown_event'` to maintain data consistency. Next, we grouped the data by key facility attributes such as `facility_id`, `facility_name`, `suburb`, `postcode`, `latitude`, and `longitude`, and aggregated the different sports offered at each facility. This involved concatenating unique values in the `sports_played` column, ensuring that all activities available at a given facility were represented in a single entry. These steps resulted in a clean, consistent dataset ready for integration and further analysis.

3. CRIME DATA

```python
# calculate the sum of incidents_recorded
incidents_per_suburb_per_year = crime_data.groupby(['year', 'suburb', 'postcode'])['incidents_recorded'].sum().r

print(incidents_per_suburb_per_year.head(20))

```

```
   year           suburb  postcode  incidents_recorded
0  2014         Abbeyard      3737                   3
1  2014       Abbotsford      3067                 889
2  2014       Aberfeldie      3040                 132
3  2014        Aberfeldy      3825                   2
4  2014          Acheron      3714                   3
5  2014            Agnes      3962                   1
6  2014          Aintree      3336                  10
```

In this step, we calculated the sum of incidents recorded per suburb, per year, by grouping the data based on year, suburb, and postcode. The total incidents for each combination were then summed up to provide a clear view of the crime trends in different areas over time. Additionally, we consolidated the year and incidents_recorded columns into a single field named incidents_records_2014_2023, creating a more streamlined and informative dataset for further analysis.

4. PLAYGROUND DATA

```
# split the geo_point
pg_data[['latitude', 'longitude']] = pg_data['geo_point_2d'].str.split(',', expand=True)

pg_data['latitude'] = pg_data['latitude'].str.strip().astype(float)
pg_data['longitude'] = pg_data['longitude'].str.strip().astype(float)

```

```
changed_pg_data = pg_data[['latitude', 'longitude', 'name']]

print(changed_pg_data.head())
```

```
    latitude   longitude                                        name
0 -37.797887  144.925942                       Holland Park Playground
1 -37.820154  144.944761                     Victoria Green Playground
2 -37.838794  144.985052                       Fawkner Park North East
3 -37.781222  144.962143                  Princes Park North Playground
4 -37.796615  144.920605  Mercantile Parade and Tankard Street Playground
```

In this data preparation step, we transformed the `geo_point_2d` field, which contained combined latitude and longitude data, into separate `latitude` and `longitude` columns. We achieved this by splitting the `geo_point_2d` field into its constituent parts and converting these values into float type for accurate geographical representation. The final dataset includes the separated `latitude`, `longitude`, and the associated playground `name`, ensuring the data is ready for geospatial analysis and mapping.

5. LANDMARK DATA

```
filtered_lm_data[['latitude', 'longitude']] = filtered_lm_data['co_ordinates'].str.split(',', expand=True)

filtered_lm_data['latitude'] = filtered_lm_data['latitude'].str.strip().astype(float)
filtered_lm_data['longitude'] = filtered_lm_data['longitude'].str.strip().astype(float)

```

In this step, we processed the co_ordinates field, which contained latitude and longitude data as a combined string. We split this field into separate latitude and longitude columns for better clarity and usability. After splitting, the data was cleaned by stripping any extra spaces and converting the values into float type to ensure they were properly formatted for geographic analysis. This transformation makes the data ready for accurate mapping and spatial queries.

3.2 Data Transformation:

- We mapped datasets with full geographic information, such as landmark, facility, and playground, to a central Location table, using fields like postcode, latitude, and longitude. For crime and accident datasets, which only had postcode and suburb data, we created a separate Suburbs_Info table to maintain a clean and organized structure, avoiding null values.

```python
# Count the number of records in each table
fc_count = filtered_fc_data.groupby(['postcode', 'suburb']).cumcount().max() + 1
lm_count = filtered_lm_data.groupby(['postcode', 'suburb']).cumcount().max() + 1
pg_count = filtered_pg_data.groupby(['postcode', 'suburb']).cumcount().max() + 1

# generate corresponding match_order columns in each table
filtered_fc_data['match_order'] = filtered_fc_data.groupby(['postcode', 'suburb']).cumcount() + 1
filtered_lm_data['match_order'] = filtered_lm_data.groupby(['postcode', 'suburb']).cumcount() + fc_count + 1
filtered_pg_data['match_order'] = filtered_pg_data.groupby(['postcode', 'suburb']).cumcount() + fc_count + lm_co
filtered_accident_data['match_order'] = filtered_accident_data.groupby(['postcode', 'suburb']).cumcount() + fc_c
filtered_crime_data['match_order'] = filtered_crime_data.groupby(['postcode', 'suburb']).cumcount() + fc_count +

# Combine postcode, suburb, latitude, longitude from all tables, keeping all combinations
location_data = pd.concat([
    filtered_fc_data[['postcode', 'suburb', 'latitude', 'longitude', 'match_order']],
    filtered_lm_data[['postcode', 'suburb', 'latitude', 'longitude', 'match_order']],
    filtered_pg_data[['postcode', 'suburb', 'latitude', 'longitude', 'match_order']],
    filtered_accident_data[['postcode', 'suburb', 'match_order']],
    filtered_crime_data[['postcode', 'suburb', 'match_order']]
], ignore_index=True)

# Generate a unique location_id for each combination
location_data['location_id'] = range(1, len(location_data) + 1)

# Merge location_ids back into individual tables
filtered_fc_data = filtered_fc_data.merge(location_data, on=['postcode', 'suburb', 'latitude', 'longitude', 'mat
filtered_lm_data = filtered_lm_data.merge(location_data, on=['postcode', 'suburb', 'latitude', 'longitude', 'mat
filtered_pg_data = filtered_pg_data.merge(location_data, on=['postcode', 'suburb', 'latitude', 'longitude', 'mat
```

4. Image Recognition Model

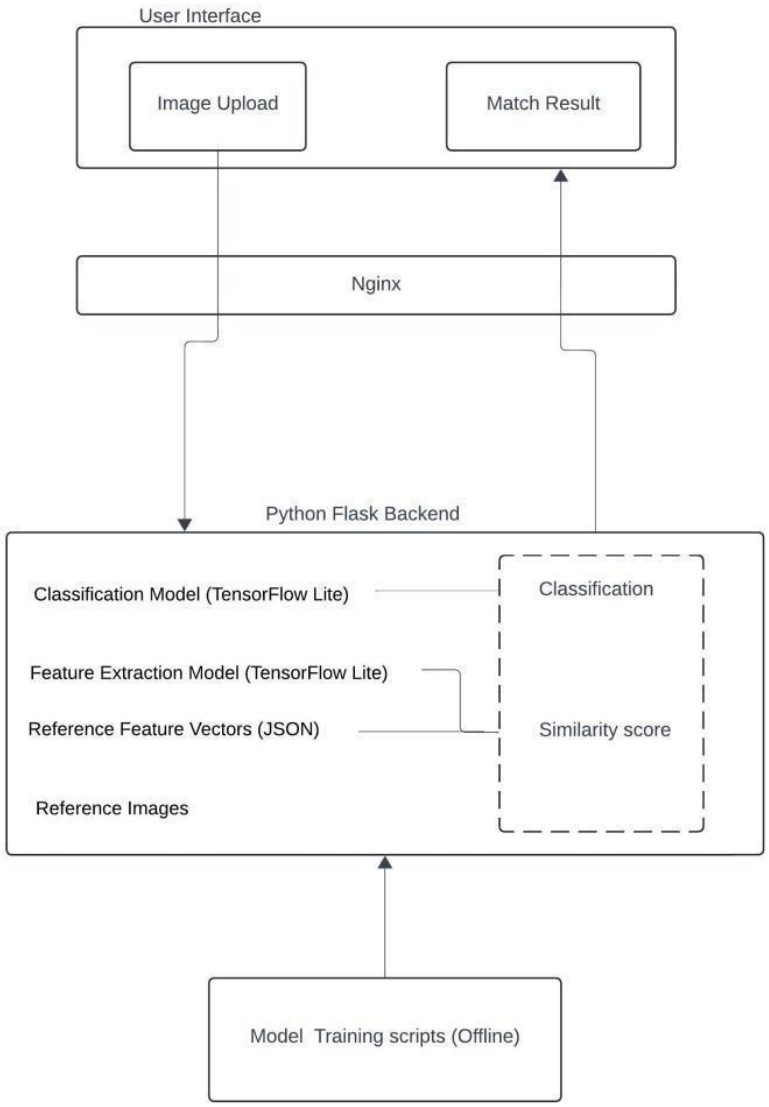
For the Scavenger Hunt feature in our platform, we implemented a machine learning-based image recognition model to compare user-uploaded images with reference photos of parks and landmarks taken by our team. This model plays a key role in verifying whether a user has correctly identified a location in the scavenger hunt, enhancing the interactive gameplay experience.

Model Design:

We used two models for the image recognition system:

1. **Classification Model:**
    - **Architecture:** A transfer learning technique was used, based on the MobileNetV2 architecture, which was pre-trained on the ImageNet dataset. The top layers of MobileNetV2 were replaced with a custom classification head, consisting of a global average pooling layer followed by two dense layers. The final layer uses softmax activation to predict one of the predefined classes (e.g., different parks or landmarks).
    - **Training:** The classification model's top layers were trained using a resized and augmented dataset, while the base layers of MobileNetV2 were frozen to retain pre-learned features. The model was trained for 10 epochs with a batch size of 32, using an 80/20 split between training and validation data. Accuracy was used as the primary performance metric.
    - **Deployment:** The classification_interpreter uses a TFLite model to categorize uploaded images into pre-defined classes, predicting the most likely location.
2. **Feature Extraction Model:**

- **Architecture:** This model was built by using the same MobileNetV2 architecture, but the classification layers were removed, retaining only the base model's output. The feature extraction model generates a feature vector for each image.
- **Similarity Comparison:** The extracted feature vector is compared to precomputed vectors from reference images using cosine similarity, allowing us to identify the closest matching park or landmark.



Key Functions:

- **invoke_model:** A helper function that feeds input tensors into the models and retrieves output tensors.

Prediction Flow:

- After the image is preprocessed, it is passed through both the classification and feature extraction models.
- The system predicts the class (location) with the highest probability and compares it against reference images using similarity scores from the feature extraction model.

Model Output:

The model returns a JSON response, including the predicted class (e.g., park name), probabilities for each class, and similarity scores. This allows the system to validate if the user has correctly identified the location in the scavenger hunt.

5. Data Storage

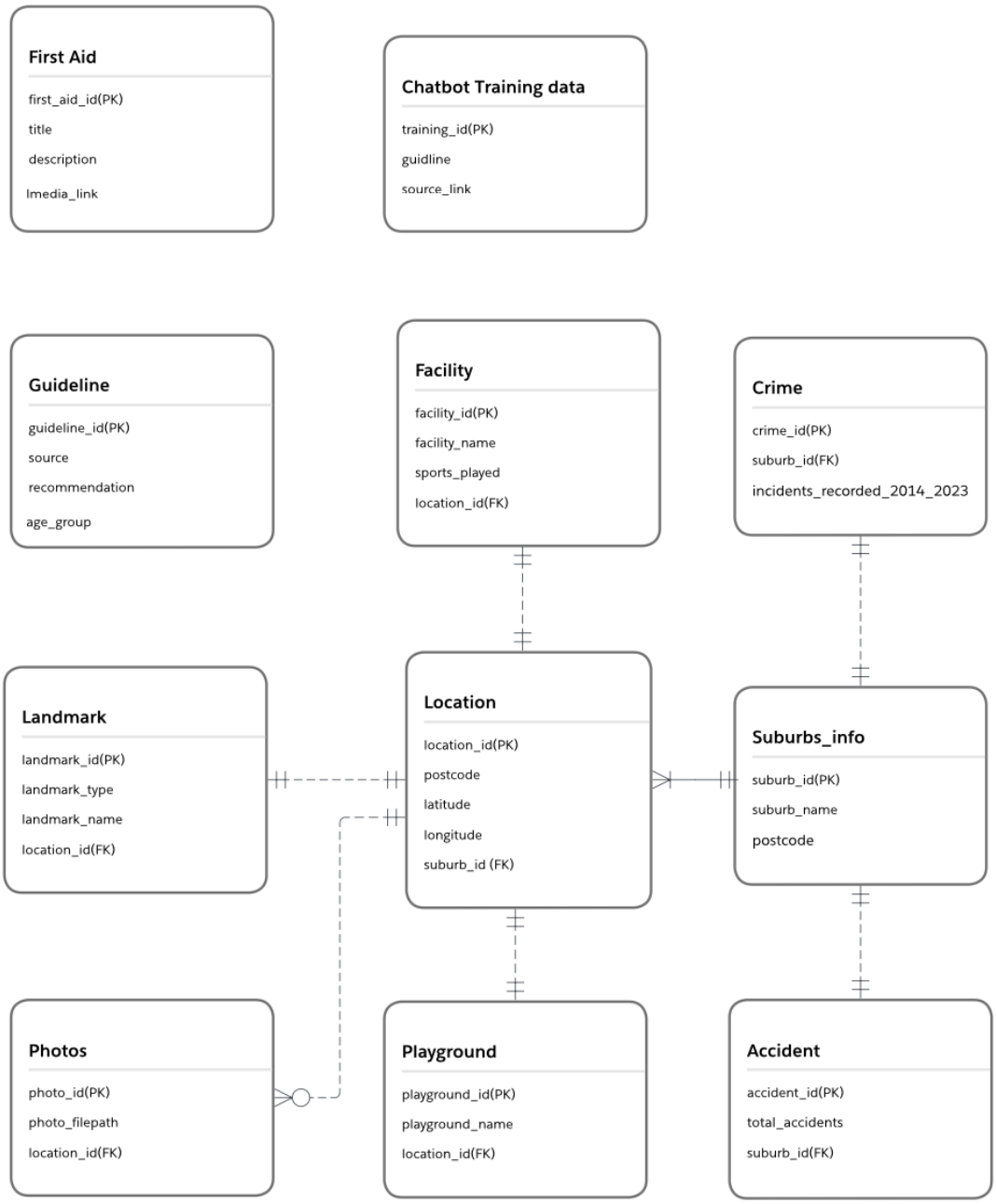
To ensure the security and integrity of our datasets, we employ a dual approach to data storage. The processed data is securely stored using Amazon Web Services (AWS) for cloud-based operations and SQLite for local development and testing.

```
# Create SQLite database
conn = sqlite3.connect('/Users/zhengzekai/FIT5120-Project/Database/safety_map.db')

# create cursor object
cursor = conn.cursor()

# Create Location table
cursor.execute('''
    CREATE TABLE IF NOT EXISTS Location (
        location_id INTEGER PRIMARY KEY,
        postcode TEXT,
        suburb TEXT,
        latitude REAL,
        longitude REAL
    )
''')

# Create Facility table
cursor.execute('''
    CREATE TABLE IF NOT EXISTS Facility (
        facility_id TEXT PRIMARY KEY,
        facility_name TEXT,
        sports_played TEXT,
        location_id INTEGER,
        FOREIGN KEY (location_id) REFERENCES Location(location_id)
    )
''')

# Create landmark table
cursor.execute('''
    CREATE TABLE IF NOT EXISTS Landmark (
        landmark_id TEXT PRIMARY KEY,
```

- **Local Development**: For local development and testing purposes, we use SQLite, a lightweight, self-contained database engine. SQLite allows for rapid prototyping and testing without the overhead of managing a server-based database.
- **Data Portability**: Data can be easily transferred between SQLite and AWS RDS, ensuring consistency across development and production environments. This approach facilitates smooth transitions from local development to cloud deployment.

6. Database Design

ER Diagram: The image below shows how the entities are connected, we have one table per each entity.



In our project, the database design is centered around the **Location** table, which serves as the central hub connecting various other tables like **Facility**, **Crime**, **Accident**, **Playground**, **Landmark**, and new tables like **Photos** and **Guideline**. This design allows us to effectively manage and query data related to different locations within Melbourne, helping us meet the goals of providing safe and accessible locations for physical activities and personalized recommendations for physical activity.

In **Iteration 3**, we extend the existing database structure with two new tables: **First Aid** and **Chatbot Training Data**. These additions enable us to deliver educational content and improve chatbot functionality, ensuring that children receive personalized, relevant guidance on safety and physical activities.

Key Entities in the Database:

1. **Location**:
    - Purpose: The Location table stores essential geographical information such as postcode, latitude, and longitude. This table is crucial as it connects other entities through the location_id, allowing us to aggregate and query data related to specific locations.
    - Key Attributes: location_id, postcode, latitude, longitude, suburb.
2. **Facility**:
    - Purpose: This table stores details about various sports and recreational facilities, helping immigrant children find locations that are safe and meet their physical activity needs.
    - Key Attributes: facility_id, facility_name,location_id, sports_played.
3. **Crime**:
    - Purpose: The Crime table records crime incidents by location, year, and type of crime. This information helps assess the safety of different areas, which is vital for guiding immigrant children to safer locations.
    - Key Attributes: crime_id, suburb_id, incidents_recorded_2014_2023.
4. **Accident**:
    - Purpose: This table logs details of accidents, including their severity and exact location. By analyzing accident data, we can identify and avoid dangerous routes, further enhancing immigrant children safety.
    - Key Attributes: accidents_id, total_accidents, suburb_id.
5. **Playground**:
    - Purpose: The Playground table contains information about playgrounds, including their names and locations. This helps immigrant children find accessible places for physical activities, especially for children.
    - Key Attributes: playground_id, playground_name, location_id.
6. **Landmark**:

- Purpose: This table stores information about various landmarks in Melbourne, providing context and aiding navigation for immigrant children.

- Key Attributes: landmark_id, landmark_type, landmark_name, location_id.

7. **Suburbs_info:**

- Purpose: This table stores information about various suburbs in Melbourne, including postcode and suburb names. It is primarily used to link datasets such as crime and accident data, which only contain suburb-level geographic information, ensuring these datasets can be integrated into the broader analysis framework.

- Key Attributes: suburb_id, postcode, suburb_name.

8. **Photos** (for Iteration 2):

- Purpose: The Photos table stores photos taken by the development team from various parks and landmarks in Melbourne. These photos are used to compare and verify user-uploaded images during the scavenger hunt. The photos are linked to specific locations to ensure accuracy in the scavenger hunt.

- Key Attributes: photo_id, photo_filepath, location_id.

9. **Guideline** (for Iteration 2):

- Purpose: The Guideline table stores structured information about various physical activity guidelines sourced from authoritative organizations like the Australian Government, WHO, and AAP. These guidelines are used to provide parents with personalized recommendations regarding their children's physical activity levels.

- Key Attributes: guideline_id, source, recommendation, age_group.

10. **First Aid** (New for Iteration 3):

- Purpose: This new table stores essential first aid guidelines, which are incorporated into the app's educational content. The first aid table provides key injury prevention techniques in a child-friendly format.

- Key Attributes: first_aid_id, title, description, media_link.

11. **Chatbot Training Data** (New for Iteration 3):

- Purpose: Stores data to train the chatbot with question-answer pairs or relevant guidelines, ensuring the chatbot can provide accurate responses for both physical activities and website navigation.

- Key Attributes: training_id, guideline, source_link.

Design Approach:

- Our database design follows the third normal form (3NF), ensuring that each table represents a unique relationship with no redundant data.

Design Rationale:

- **Centralised Location Data**: The **Location** and **Suburbs_info** tables continue to serve as core reference points for all geographically-based data. This structure remains essential for cross-referencing entities like **Facility**, **Playground**, **Landmark**, and **Photos**. The addition of the **First Aid** and **Chatbot Training Data** tables extends the ability to offer personalized safety advice and chatbot assistance without disrupting existing functionality.

- **Foreign Key Relationships**: Each entity remains linked to either the **Location** or **Suburbs_info** tables via location_id or suburb_id, ensuring data consistency and accuracy. The new **First Aid** and **Chatbot Training Data** tables operate independently but integrate seamlessly into the broader data structure by contributing to both the educational content and AI-based guidance.

- **Scalability**: The database design remains flexible and scalable, allowing the addition of new datasets and tables as needed. The introduction of the **First Aid** and **Chatbot Training Data** tables in Iteration 3 demonstrates the system's capacity to evolve, supporting the growing complexity of features like first aid education and personalized chatbot recommendations.

7. Data Analytics

For each dataset, we have analyzed its potential for generating **insights**, **foresights**, and **hindsights** to guide decisions and improve the physical activity of immigrant children.

7.1 Iteration 1

Sport and Recreational Facilities List:

- **Insight**: We analyze the current distribution of sports and recreational facilities across Melbourne to identify well-serviced areas versus those lacking sufficient resources. This will help ensure that immigrant children have easy access to facilities, making it more convenient for them to engage in physical activities.
- **Foresight**: By evaluating demographic trends and growth patterns, we can predict future needs for facilities in expanding neighborhoods. This allows us to recommend the strategic placement of new facilities in areas where they will be most needed, helping to plan for the future well-being of children.
- **Hindsight**: Reviewing the historical placement and usage of facilities helps us understand past gaps in accessibility and how they have been addressed. This context helps ensure that current recommendations for facility improvements are grounded in an understanding of long-term community needs.

Crime Incident Dataset:

- **Insight**: By combining crime data with accident and weather conditions, we can calculate safety rates for different locations. This data is used to provide real-time safety assessments to immigrant children, helping them avoid unsafe areas when planning physical activities.
- **Foresight**: Analyzing crime trends over the last decade allows us to predict potential future crime hotspots, enabling us to proactively guide immigrant children away from areas that may become unsafe in the near future.
- **Hindsight**: Examining historical crime data helps us identify long-standing safety concerns in specific areas, providing a comprehensive view of how safety has evolved in Melbourne. This information is crucial for making informed decisions about current and future safety recommendations.

Playground Dataset:

- **Insight**: We analyze the accessibility and condition of playgrounds to provide immigrant children with recommendations for nearby, safe, and well-maintained playgrounds. This ensures that children have access to quality spaces for physical activity.
- **Foresight**: By comparing playground data with demographic information, we can anticipate areas where new playgrounds will be needed as the population of children grows, allowing for proactive planning and resource allocation.
- **Hindsight**: Historical data on playground usage and condition helps us understand patterns of neglect or underutilization, ensuring that current recommendations address these issues and improve access to safe play areas.

Victoria Road Crash Dataset:

- **Insight**: We will use road crash data to identify accident hotspots and categorize areas by risk level. This allows us to recommend safer routes for physical activities, particularly for children who need to avoid dangerous locations.
- **Foresight**: By analyzing accident trends, we can predict where future accidents are likely to occur, helping us guide immigrant children toward safer paths before issues arise, thus preventing potential accidents.
- **Hindsight**: Reviewing past accident data gives us a detailed understanding of where and why accidents have occurred. This historical perspective is essential for making informed recommendations that enhance road safety for immigrant children.

Landmark Dataset:

- **Insight**: We analyze landmark data to create visual references on the map, aiding navigation and helping immigrant children find safe and accessible locations for physical activities. This enhances the overall immigrant children experience by providing familiar points of reference.
- **Foresight**: Analyzing trends in the popularity and usage of landmarks helps us anticipate future changes in how these areas are used, allowing the map to stay relevant and useful as immigrant children need to evolve.

- **Hindsight**: Understanding the historical significance and changes in landmark usage provides context for their current importance, ensuring that the map reflects accurate and meaningful information for immigrant children.

7.2 Iteration 2

Australian Government's Physical Activity Guidelines for Children and Young People:

- **Insight:** We analyze the official guidelines to provide structured recommendations for children's physical activity based on age and developmental stages. This ensures that the recommendations in our app are aligned with the most current health standards in Australia.
- **Foresight:** By tracking trends in physical activity levels among children, we can predict which age groups or activities may need more attention in the future. This allows us to proactively recommend activities that will help children meet their daily physical activity targets.
- **Hindsight:** Reviewing past adherence to the guidelines helps us understand how effectively children have been meeting the recommended physical activity levels. This information guides us in refining our recommendations and identifying areas for improvement in promoting healthier activity habits.

Guideline Table (combined source data for parental guidance):

- **Insight:** We analyze the guideline data in combination with **survey responses** from parents, considering factors such as **age, current physical activity levels**, and **screen time** to provide personalized recommendations for improving children's physical activity."
- **Foresight:** By tracking responses to the guidelines, such as increased **outdoor playtime** or **reduced screen time**, we can anticipate future trends in physical activity and adjust our recommendations to encourage healthier behavior patterns.
- **Hindsight:** We will review **historical adherence** to the guidelines based on periodic **survey feedback** from parents, allowing us to measure the effectiveness of our recommendations and refine them to ensure children are meeting physical activity targets.

7.3 Iteration 3

First Aid and Safety Guidelines for Children & Australian Red Cross First Aid for Children:

- **Insight:** We draw from both Health.gov.au and the Australian Red Cross to provide children with practical and accessible first aid information. By combining these resources, we ensure that children have access to comprehensive guidelines on basic injury prevention and first aid practices tailored for outdoor activities. These guidelines are crucial in helping children stay safe while engaging in physical activities in parks and other outdoor environments.
- **Foresight:** By analyzing trends in childhood injuries during outdoor play, we can predict which first aid techniques will be most important for the future. As more children participate in outdoor activities, we expect an increased need for knowledge

on treating minor injuries, such as falls, cuts, and bruises. We are preparing the app to highlight these techniques based on anticipated future needs.

- **Hindsight:** We look back at historical data on common injuries and the effectiveness of past first aid interventions to understand which injuries are most frequent in outdoor settings. By doing so, we refine our educational content to ensure it addresses the most common safety challenges that children face during physical activities, allowing us to provide relevant and practical first aid advice.

8. Ethical, Legal and Privacy Issues

In developing the platform to help immigrant children in Melbourne find safe and accessible locations for physical activity, as well as offering educational content on first aid and safety, upholding the highest standards of ethics, legality, and privacy is paramount. This commitment influences every aspect of our data management practices, from the selection of datasets to the deployment of analytics and chatbot interactions.

8.1 Ethical Considerations:

Our project is built on a foundation of ethical guidelines that ensure respect for individual privacy and data integrity. We are committed to:

- **Handling Uploaded Photos**: Photos taken by users during the scavenger hunt are only used for real-time verification and are not stored permanently. This ensures that no identifiable data or images are retained, protecting the privacy of all participants.
- **AI Chatbot**: The chatbot is designed to provide safety and first aid guidance without collecting or storing personal information. It interacts with children solely based on predefined questions and guidelines, respecting their privacy.
- **Educational Content**: The educational first aid content provided through the app does not require any personal data from users, ensuring privacy is maintained while offering safety-related resources.

8.2 Legal Compliance:

Our project strictly adheres to all applicable laws and regulations, including:

- **Data Protection Laws**: We comply with the Australian Privacy Principles (APPs) to ensure the rights of data subjects are protected at all times. This includes transparent data handling practices and secure storage methods.
- **License Adherence**: All datasets used in the project, including physical activity guidelines, are governed under appropriate licenses (e.g., Creative Commons Attribution 4.0 International). We ensure proper attribution is given to all data sources, and that the data is used in accordance with the license conditions.

8.3 Privacy Protection:

We employ comprehensive security and privacy measures to ensure data integrity, including:

- **Security Measures**: We use encryption, secure data storage, and stringent access control mechanisms to prevent unauthorized access to our databases. AWS RDS's advanced security features further enhance data protection.

- **Handling Team and User-Uploaded Photos**: Photos taken by the development team for the scavenger hunt and photos uploaded by users are handled securely and used only for real-time location verification. No personal data or photos are stored after verification, and all photo data is processed in a transient manner to protect privacy.

- **Absence of User Data Collection**: Our application respects immigrant children's privacy by not capturing or storing personal data. Immigrant children do not need to enter any personal information when using our platform, ensuring their privacy is safeguarded.

- **Guidelines and Recommendations**: Physical activity and first aid guidelines are sourced from government and international organizations. Recommendations provided to users are based on survey responses and publicly available guidelines, with no personal data being collected or stored.