

Metadata and Data Dictionary

Paris for kids

By Xinly ROY

November 2024

This document provides a detailed overview of all datasets used in the Kids-Friendly Places project. It explains the structure, purpose, and attributes of each dataset, ensuring transparency and reproducibility of the analysis.

Dataset Name: detailed_places

- **Source:** Google Places API
- **Description:** Contains detailed information about officially categorized kid-friendly places in Paris.
- **Columns:**
 1. **name (string):** The name of the location (e.g., restaurant, park).
 2. **type (string):** The type of location (e.g., Restaurant, Coffee Shop).
 3. **rating (float):** The average user rating of the location.
 4. **vicinity (string):** A short address or neighborhood description.
 5. **total_ratings (integer):** The total number of user ratings.
 6. **formatted_address (string):** The full address of the location.
 7. **website (string):** The official website of the location (if available).
 8. **phone_number (string):** The phone number of the location (if available).
 9. **opening_hours (text):** Description of the opening hours.
 10. **reviews (text):** User reviews or comments about the location.
 11. **zipcode (integer):** The postal code of the location.
 12. **unique_id (string):** A unique identifier for the location.
 13. **type_id (string):** A unique identifier for the type of place.
 14. **df_id (string):** Dataset-specific unique identifier.
 15. **df_unique_id (string):** A concatenation of dataset ID and unique ID.

Dataset Name: faire_events

- **Source:** Kaggle
- **Description:** Contains data about family-friendly events in Paris.
- **Columns:**
 1. **unique_id (integer):** A unique identifier for the event.
 2. **url (string):** Link to the event's details.
 3. **titre (string):** Title of the event.
 4. **description (text):** Description of the event.
 5. **nom_lieu (string):** Name of the event's venue.
 6. **adresse_lieu (string):** Address of the venue.
 7. **zipcode (string):** Postal code of the event's venue.
 8. **ville (string):** City where the event takes place.
 9. **coordonnees_geographiques (string):** Geographic coordinates of the venue.
 10. **type_prix (string):** Price type (e.g., free, paid).
 11. **type_acces (string):** Access type (e.g., recommended, mandatory).
 12. **audience (string):** Target audience of the event.
 13. **locale (string):** Locale or language of the event.
 14. **month_debut (string):** Month the event starts.
 15. **year_debut (integer):** Year the event starts.
 16. **month_fin (string):** Month the event ends.
 17. **year_fin (integer):** Year the event ends.
 18. **type (string):** Type of event (e.g., cultural, educational).
 19. **type_id (string):** A unique identifier for the type of event.
 20. **df_id (string):** Dataset-specific unique identifier.
 21. **df_unique_id (string):** Concatenation of dataset ID and unique ID.

Dataset Name: paris_data

- **Source:** Web Scraping (Wikipedia, Paris.fr)
 - **Description:** Contains demographic and geographic data for Paris.
 - **Columns:**
 1. **arrondissement (integer):** District number of Paris.
 2. **name_arrondissement (string):** Official name of the arrondissement.
 3. **surface_ha (float):** Surface area of the arrondissement in hectares.
 4. **population_2020 (integer):** Population of the arrondissement in 2020.
 5. **density_2021_hab_km2 (integer):** Population density in 2021.
 6. **zipcode (string):** Postal code of the arrondissement.
-

Dataset Name: parks

- **Source:** CSV (Opendata.paris)
- **Description:** Contains information about parks and playgrounds in Paris.
- **Columns:**
 1. **unique_id (string):** Unique identifier for the park.
 2. **name_park (string):** Name of the park.
 3. **typologie_espace_vert (string):** Type of green space.
 4. **categorie (string):** Category of the park.
 5. **zipcode (string):** Postal code of the park.
 6. **ouverture_24h_24h (string):** Indicates whether the park is open 24/7.
 7. **url_plan (string):** Link to the park's layout or plan.
 8. **geo_point (string):** Geographic coordinates of the park.
 9. **adresse (string):** Address of the park.
 10. **type (string):** Type of park (e.g., recreational, nature).
 11. **type_id (string):** Unique identifier for the type of park.
 12. **df_id (string):** Dataset-specific unique identifier.
 13. **df_unique_id (string):** Concatenation of dataset ID and unique ID.

Dataset Name: type

- **Source:** Manually Created
 - **Description:** Centralized type table connecting all datasets through shared type categories.
 - **Columns:**
 1. **type_id (string):** Unique identifier for the type.
 2. **type (string):** Description of the type (e.g., Restaurant, Park).
 3. **zipcode (string):** Associated zip code for the type.
-

Dataset Name: centralized_types

- **Source:** Consolidated Table
- **Description:** Combines key data from all datasets into a single table for streamlined analysis.
- **Columns:**
 1. **unique_id (string):** Unique identifier across all datasets.
 2. **type_id (string):** Unique identifier for the type.
 3. **type (string):** Description of the type.
 4. **name (string):** Name of the entity (e.g., park name, event title).
 5. **zipcode (string):** Postal code of the entity.
 6. **arrondissement (string):** District of the entity.