Task B2: Discover and plan: library checkouts

Data Questions:

**Question 1:** What are the most borrowed item types in the library, and how has their popularity changed over the years?

* Transformation Steps: Combine data from librarycheckouts\_2014, librarycheckouts\_2015, and librarycheckouts\_2016. Group by item type and year to count checkouts.

**Question 2:** Where are the high-traffic locations in the library, and how does it vary by collection?

* Transformation Steps: Join librarycheckouts data with library\_inventory\_matches and library\_System\_Data\_Dictionary to get location details. Group by location and collection to analyse foot traffic.

**Question 3:** Who are the top borrowers, and what types of items do they frequently check out?

* Transformation Steps: Analyse librarycheckouts data to identify users with the highest checkout counts. Join with library\_inventory\_matches for item details.

**Question 4:** Are there any seasonal trends in library usage? How do checkouts vary throughout the year?

* Transformation Steps: Group checkouts by month and visualize trends over multiple years

**Question 5:** How diverse is the library’s collection in terms of item types, and has this diversity changed over the years?

* Transformation Steps: Analyse the distribution of item types and track changes annually.

**Question 6:** Are there specific days of the week when the library experiences higher foot traffic?

* Transformation Steps: Group checkouts by day of the week and analyse patterns.

Transformation Steps:

1. **Combine Data:** Merge librarycheckouts\_2014, librarycheckouts\_2015, and librarycheckouts\_2016 into a single dataset for comprehensive analysis.
2. **Standardise Item Types:** Utilise the library\_System\_Data\_Dictionary to standardise item types across years for consistent analysis.
3. **Enrich Location Information:** Join librarycheckouts data with library\_inventory\_matches and library\_System\_Data\_Dictionary to enrich location details.
4. **User Analysis:** Aggregate librarycheckouts data to identify top borrowers based on checkout counts.
5. **Temporal Analysis:** Consider creating a date hierarchy for time-based analysis, enabling comparisons over months and years.
6. **Data Quality Checks:** Check for missing or inconsistent values and handle them appropriately for accurate insights.
7. **Data Visualisation:** Plan charts/graphs to effectively communicate findings for each data question.

#### **Additional Notes from CSVs:** Call number: tells us the address of the book in the library, where it is located.

Bib number: a unique identifier for each bibliographic record in the catalog.

Item type: