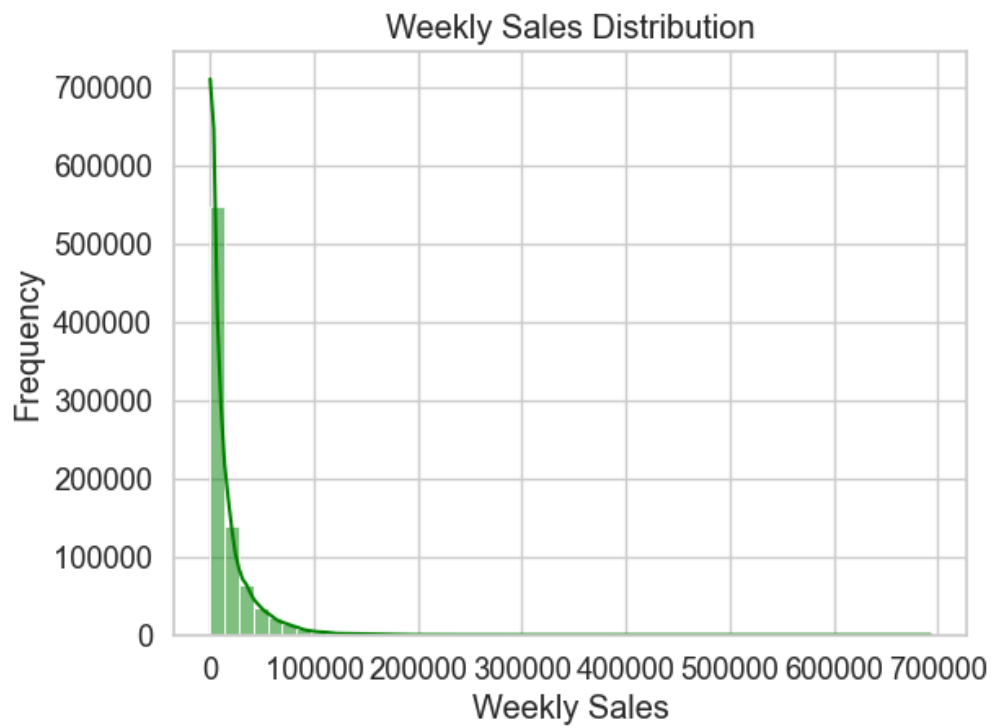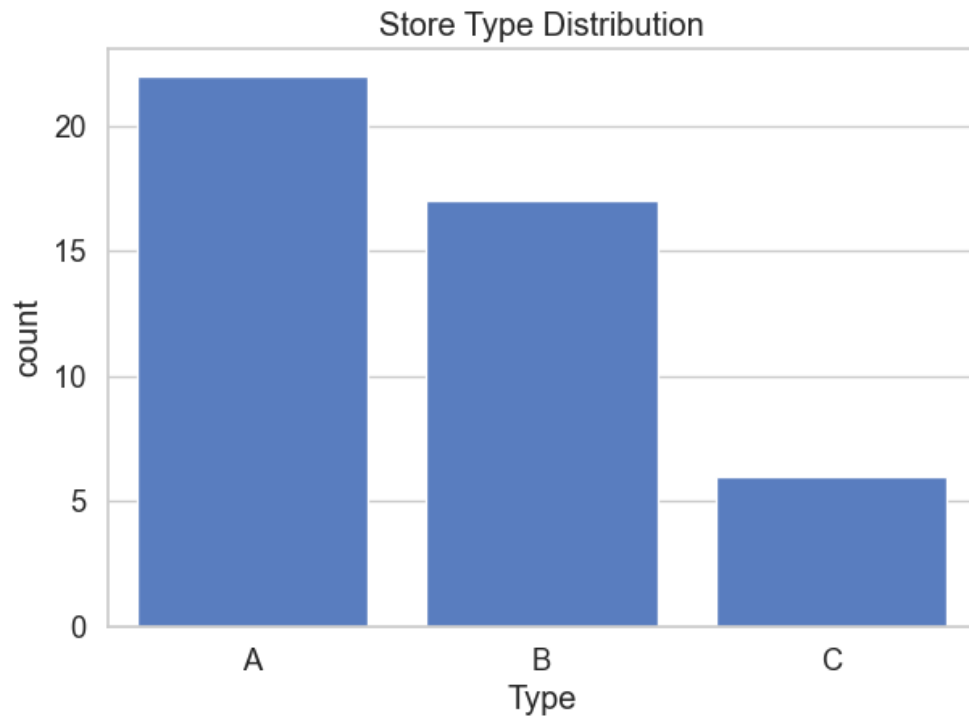# 1. Introduction

- Goal: Sales (Retail), Stores (Retail), Features (Retail).

- Purpose To analyze the data-quality dimensions (complete, valid, unique, consistent) and develop a star schema to warehouse.

- Significance: Real time ETL means good analytics and decision making.

2. Data Quality Analysis

As stated in your ETL results you have the following evidence:

- **Dim Store evidence**: After deduplication, unique stores have surrogate keys assigned to them.

- **Dim Date evidence**: All the sales dates with Year, Month, Week, Quarter attributes added to it.

- **Dim Features evidence**: The combination of Store Date was validated, and duplicates were professional ed out.

- **Fact Sales records**: 840,424 records which survived the cleaning process (positive Weekly Sales, valid Store Date).

- **Orphan keys evidence**: Number of sales using missing stores/dates of "fk_integrity_fact.csv".

- **Unknown store ratio evidence**: Ratio of sales associated with surrogate "Unknown store.

- **Missing values evidence**: Completeness problems revealed in raw data.

- **Duplicates evidence**: Duplicate records indicated in uniqueness check.

Store Type Distribution

Weekly Sales Distribution

3. Recommendations

- **Surrogate processing**: Expand orphan key dimensions with Unknowns.

- **Rules of ETL validation**: Positive Weekly Sales, valid date and Deduction.

- **Governance:** Keep records of missing values, duplication of records and names of orphan records that will be tracked.

- **Schema design**: Make sure that there is conforming dimensions and referential integrity among fact tables.

4. Conclusion

- The process of ETL pipeline was able to process raw data into clean stars schema.

- Quality of data was found and addressed.

- There is evidence of better consistency, validity and completeness.

- Recommendations are in line with the concepts of weekly 10 (data quality dimensions, referential integrity, governance).

5. References

- Dataset source (Retail Sales, Stores, Features).

- Course notes and textbook.

- AI assistance (Microsoft Copilot).