Datasets from IBM BI Analyst Certification on Coursera.

Excel Project

# Data Source

sales.csv - <https://bi-analyst-capstone-sk2280913-83ef34ea5b1e2138c3b6e6f9dff6105f1.gitlab.io/labs/Module1/Lesson%201_Data/sales.csv>

product\_hierarchy.csv - <https://bi-analyst-capstone-sk2280913-83ef34ea5b1e2138c3b6e6f9dff6105f1.gitlab.io/labs/Module1/Lesson%201_Data/product_hierarchy.csv>

store\_cities.csv - <https://bi-analyst-capstone-sk2280913-83ef34ea5b1e2138c3b6e6f9dff6105f1.gitlab.io/labs/Module1/Lesson%201_Data/store_cities.csv>

store\_names.csv - <https://bi-analyst-capstone-sk2280913-83ef34ea5b1e2138c3b6e6f9dff6105f1.gitlab.io/labs/Module1/Lesson%201_Data/store_names.csv>

city\_names.csv - <https://bi-analyst-capstone-sk2280913-83ef34ea5b1e2138c3b6e6f9dff6105f1.gitlab.io/labs/Module1/Lesson%201_Data/city_names.csv>

product\_names - <https://bi-analyst-capstone-sk2280913-83ef34ea5b1e2138c3b6e6f9dff6105f1.gitlab.io/labs/Module1/Lesson%201_Data/product_names.csv>

# Data Schema

sales.csv - 13 Columns, 49290 Rows

* Columns:
  + Product\_id
  + Store\_id
  + Date
  + Sales
  + Revenue
  + Stock
  + Price
  + Promo\_type\_1
  + Promo\_bin\_1
  + Promo\_type\_2
  + Promo\_bin\_2
  + Promo\_discount\_2
  + Promo\_discount\_type\_2

product\_hierarchy.csv - 10 Columns, 634 Rows

* Columns:
  + Product\_id
  + Product\_length
  + Product\_depth
  + Product\_width
  + Cluster\_id
  + Hierarchy1\_id
  + Hierarchy2\_id
  + Hierarchy3\_id
  + Hierarchy4\_id
  + Hierarchy5\_id

store\_cities.csv - 4 Columns, 145 Rows

* Columns
  + Store\_id
  + Storetype
  + Store\_size
  + City\_id

store\_names.csv - 2 Columns, 145 Rows

* Columns
  + Store\_id
  + Store\_name

city\_names.csv - 2 Columns, 38 Rows

* Columns
  + City\_id
  + City\_name

product\_names - 2 Columns, 700 Rows

* Columns
  + Product\_id
  + Product\_name

# Data Extraction

1. Open a new Excel workbook.
2. Save as “excel\_project” .
3. Go to Data Tab and load “city\_names.csv” in its respective sheet.
4. Go to Data Tab and load “product\_hierarchy.csv” in its respective sheet, without the “cluster\_id”, ”hierarchy1\_id”, ”hierarchy2\_id”, ”hierarchy3\_id”, ”hierarchy4\_id”, and ”hierarchy5\_id” columns.
5. Go to Data Tab and load “product\_names.csv” in its respective sheet.
6. Go to Data Tab and load “sales.csv” in its respective sheet, without the “promo\_type\_1”, “promo\_bin\_1”, “promo\_type\_2”, “promo\_bin\_2”, “promo\_discount\_2” and “promo\_discount\_type\_2” columns.
7. Go to Data Tab and load “store\_cities.csv” in its respective sheet.
8. Go to Data Tab and load “store\_names.csv” in its respective sheet.

# Data Aggregation

1. Perform Merge Queries between “sales” sheet and “product\_hierarchy” sheet with JOIN KIND – Left Outer (all from first, matching from second) based on column “product\_id”.
2. On the previous merge we merge further the “product\_names” sheet with JOIN KIND – Left Outer (all from first, matching from second) based on column “product\_id”.
3. On the previous merge we merge further the “store\_names” sheet with JOIN KIND – Left Outer (all from first, matching from second) based on column “store\_id”.
4. On the previous merge we merge further the “store\_cities” sheet with JOIN KIND – Left Outer (all from first, matching from second) based on column “store\_id”.
5. On the previous merge we merge further the “city\_names” sheet with JOIN KIND – Left Outer (all from first, matching from second) based on column “city\_id”.

The result: 18 Columns and 49289 Rows.

# Data Cleaning

1. Spell Checking.
2. Assign the right data formats.

# Working Copy

1. Create a “working\_copy” sheet from “merged\_data” sheet.
2. Freeze top row.
3. Create new columns in PowerQuery to extract Day Name, Day Number, Month Name and Year from date.
4. Remove unnecessary columns.
5. Creating a pivot tables for range “A1:M49290” with their respective visualisations.