Objective

This document serves as a guide to understanding the Annie's Magic Numbers use case. It outlines the initial approaches, exploratory data analysis, and diagrams, providing a structured overview of the process.

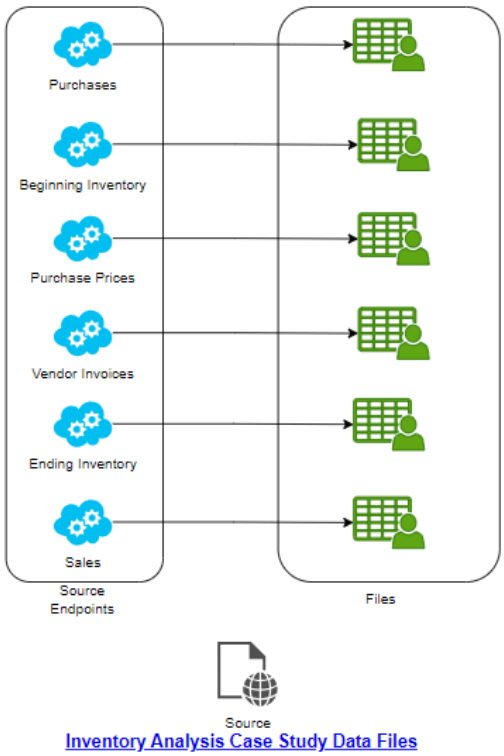
Data Sources

Location of the data

The data sources come from an endpoint available on the [Data analytics case study data files] page. Each endpoint provides CSV files that must be ingested to extract the information required by stakeholders.

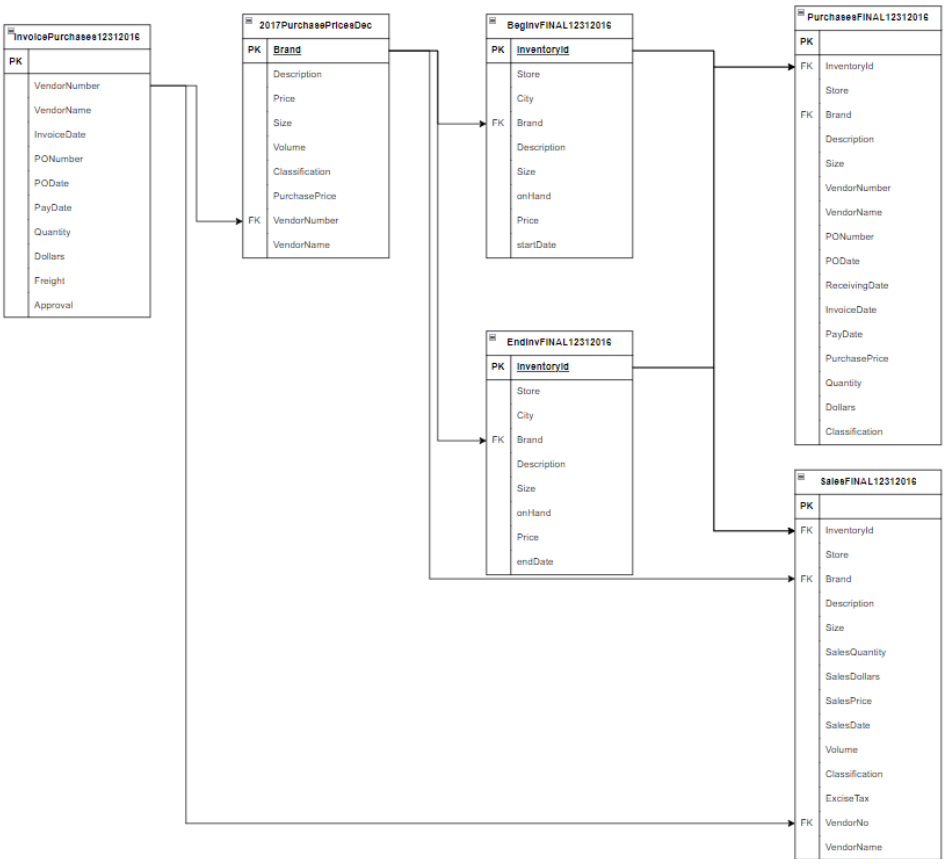
Data Relationships

In the initial analysis of the source, we identified data extracted from a transactional system. Some of these datasets appear to be related, as illustrated in the following diagram:



Some tables do not have a primary key, so it will be necessary to create one in the Data Warehouse.

During the initial analysis, we identified that some data is extracted from a transactional source. Based on this, we can infer potential relationships between datasets, as illustrated in the following diagram:



Calculations required

The calculation requested by the stakeholders are the following: **profits** and **margins**

Profit

$$\text{Profit} = \text{Total Revenue} - \text{Total Cost}$$

Margin

$$\text{Profit Margin} = \left(\frac{\text{Profit}}{\text{Total Revenue}} \right) \times 100$$

Exploratory Data Analysis
To perform the initial exploratory data analysis (EDA), we manually loaded the information into the Databricks Unity Catalog (Filestore). This allowed us to implement the first version of the profit and margin calculations.

The primary tables used for this analysis are:

- Sales
- Purchase

The SQL implementation is as follows:

```sql
%sql
SELECT
    s.InventoryId,
    s.Store,
    s.SalesDescription,
    s.SalesPrice,
    p.PurchasePrice,
    (s.SalesPrice - p.PurchasePrice) AS Profit,
    ((s.SalesPrice - p.PurchasePrice) / s.SalesPrice) * 100 AS ProfitMargin
FROM
    Sales s
LEFT JOIN
    PurchasePrices p
    ON s.Brand = p.Brand
LIMIT 5;
```

The complete exploratory data analysis is available in the following Databricks notebook: [AMN - Databricks Notebook]