

# **Power BI Dataset Assessment Form**

Name : Aswin Santhosh

Project Title : **Nike Sales Analytics Dashboard**

## **Dataset Source**

Nike Sales Dataset (CSV – Uncleaned, later cleaned and normalized)

Dataset Link : <https://www.kaggle.com/datasets/nayakganesh007/nike-sales-uncleaned-dataset>

## **1. Project Goal**

The goal of this project is to analyze Nike sales data to understand revenue, profit, product performance, regional trends, and channel effectiveness. Using Power BI, the project transforms raw sales data into a structured star schema model and delivers an interactive dashboard that supports data-driven business decisions.

## **2. Dataset Overview**

### **Source of Dataset**

- Internal / Academic Nike Sales Dataset

### **Data Format**

- CSV (Comma-Separated Values)

### **Dataset Nature**

- Historical sales data
- Transaction-level dataset

## 3.Data Quality

### **How will missing or invalid values be handled?**

Missing and invalid values are handled during the data cleaning stage using Power BI Power Query. Records with critical missing fields (such as revenue, product ID, or date) are removed, while non-critical missing values are treated using logical defaults or exclusions to maintain data integrity.

### **Are the columns consistent in data type?**

Yes. All columns maintain consistent data types such as numeric, text, and date formats after cleaning and transformation.

### **Are there any calculated fields in the dataset?**

Yes. Several calculated fields and measures are created in Power BI using DAX, including: - Total Revenue - Total Cost - Total Profit - Profit Margin % - Average Discount

## 4.Data Structure

### Tables Used

Table Name	Type	Description
Fact_Sales	Fact Table	Stores transactional sales metrics
Dim_Date	Dimension	Date-related attributes
Dim_Product	Dimension	Product details and product lines
Dim_Region	Dimension	Regional information
Dim_Channel	Dimension	Sales channel (Online, Retail)
Dim_Gender	Dimension	Customer gender classification

### Data Model

- Star Schema
- One-to-many relationships between dimension tables and the fact table

## 5.Data Content

### Key Columns

- Revenue (Numeric)
- Units Sold (Numeric)
- Discount (Numeric)
- Average Discount (Numeric)
- Total Cost (Numeric)
- Total Profit (Numeric)
- Product Name (Text)
- Product Line (Text)
- Size (Text)
- Region (Text)
- Channel (Text)
- Date / Month / Year (Date)

## 6.Time-Based Data

**Does the dataset include time-based data? Yes.**

### Time Attributes

- Date
- Month
- Month Name
- Year

These fields are used for trend analysis, year-over-year comparison, and monthly performance tracking.

## 7.Geographical Data

**Does the dataset include geographical information? Yes.**

### Geographical Fields

- Region (Bangalore, Delhi, Hyderabad, Kolkata, Mumbai, Pune)

## 8.Numerical Ranges & Outliers

**Are there numerical fields with large ranges or potential outliers? Yes.**

Fields such as revenue, profit, cost, and discount may contain large values or variations.

### Handling Approach

Outliers are reviewed using Power BI visuals. Extreme values are validated against business logic before inclusion to ensure accuracy.

## 9.Hierarchies & Categorization

**Does the dataset include hierarchical data? Yes.**

### Hierarchies Used

- Date Hierarchy: Year → Month
- Product Hierarchy: Product Line → Product Name

## 10.Analytical Capability

**Does the dataset support trend analysis, comparison, and insights? Yes.**

The dataset allows: -

- Time-based trend analysis
- Product and product line comparisons
- Regional and channel performance evaluation
- Gender-based sales analysis
- Profitability, discount, and margin analysis

## **11.Dataset Size & Performance**

### **Approximate Dataset Size**

- Medium-sized dataset (suitable for Power BI analysis)

### **Performance Considerations**

- Star schema model improves performance
- Optimized DAX measures
- Efficient filtering and cross-highlighting

## **12.Tools & Technologies Used**

- Power BI Desktop
- Power Query (Data Cleaning & Transformation)
- DAX (Measures and Calculations)

## **13.Final Assessment**

The dataset is well-suited for analytical reporting and visualization in Power BI. After cleaning and normalization, it supports efficient querying, meaningful KPIs, and interactive dashboards aligned with business requirements.