**HEALTHCARE FABRIC POC**

**Table of Contents**

1. Problem Statement

* Overview
* Significance of Data-Driven Decision-Making in Healthcare

1. Microsoft Fabric

* Comprehensive Analytics Solution
* Data Security and Governance

1. Design/Flow

* Approach 1: Flow for Drug Data Analysis
* Approach 2: Flow for Nutrition Data Analysis

1. Naming Convention

* Drug Data
* Nutrition Data

1. Issues/Challenges
2. Step-by-Step Process

* Approach 1

6.1. Data Upload to HealthCare\_Data Folder

6.2. Conversion to Delta Tables

6.3. Creating HealthCare\_Pipeline

6.4. Cleaning and Transformations in HealthCare\_DW

6.5. Creation of Dimension and Fact Tables

6.6. Custom Views for Analysis

6.7. PowerBI Stories

* Approach 2

6.1. Loading Data to HealthCare\_Nutrition\_test\_Dataflow

6.2. Applying Transformations

6.3. Establishing HealthCare\_Test\_Pipeline

6.4. Creating Dataflow and Copy Data Activities

6.5. Dataflow Activity: Transformations and Delta Table Update

6.6. CopyData Activities: Moving Data to DW Tables

6.7. Running the Pipeline and Inserting Data

6.8. Creating Views for Analysis

6.9. PowerBI Story Creation

1. Summary

**Problem Statement**

In the dynamic healthcare landscape, data-driven decision-making is pivotal for optimizing patient care, resource allocation, and system efficiency. This analysis focuses on a comprehensive dataset containing vital information on drugs, product details, packaging specifications, and nutrition data.

This dataset is a valuable resource for healthcare professionals, administrators, and researchers aiming to enhance patient outcomes, reduce costs, and improve service delivery. Analyzing the intricate relationships within this data can drive evidence-based decisions, optimizing drug manufacturing, ensuring product safety, and tailoring nutritional interventions.

Through advanced data analysis techniques, we aim to extract meaningful patterns, correlations, and trends. The insights gained have the potential to not only improve patient care but also advance the healthcare industry.

[**Microsoft Fabric**](https://www.microsoft.com/en-us/microsoft-fabric)

Microsoft Fabric offers a comprehensive analytics solution encompassing a range of services, from data handling, data lakes, and data engineering to data integration, data science, real-time analytics, and business intelligence. This integrated platform ensures strong data security, governance, and compliance measures.

Gone are the days of piecing together disparate analytics services from various providers. Instead, opt for this seamless solution that simplifies connection, onboarding, and operation processes for your organization.

A screenshot of a computer

Description automatically generated

**Design/Flow**

* **Approach1:** Flow for Drug Data Analysis

A screenshot of a computer

Description automatically generated

* **Approach2:** Flow for Nutrition Data Analysis

A diagram of data flow

Description automatically generated

**Naming Convention -**[**Dataset**](https://www.kaggle.com/datasets/maheshdadhich/us-healthcare-data)

1. **Drug Data**

**Data location:** Local

**Dataset Names:** drugs\_product, drugs\_package, drugs\_unfinished\_product, drugs\_unfinished\_products

**Delta Table Names:** drugs\_product, drugs\_package, drugs\_unfinished\_product, drugs\_unfinished\_products

**Lakehouse Name:** HealthCare\_Lakehouse

**Pipeline Name:** HealthCare\_Pipeline

**Datawarehouse Name:** HealthCare\_DW

**Schema Table Names:** FACT\_PRODUCT, DIM\_USAGE, DIM\_PRODINFO, DIM\_MARKETING, DIM\_PACKAGE

**View Names:** vw\_FACT\_PRODUCT, vw\_DIM\_USAGE, vw\_DIM\_PRODINFO, vw\_DIM\_MARKETING,

vw\_DIM\_PACKAGE

**Data model Name:** Drugs

**PowerBI Report Name:** Demo

1. **Nutrition Data**

**Data Location:** Loca**l**

**Staging Delta Table Names:** Nutritions\_US\_Staging

**Lakehouse Name:** HealthCare\_Test\_Lakehouse

**Pipeline Name:** HealthCare\_Test\_Pipeline

**Dataflow Name:** HealthCare\_Nutrition\_test\_Dataflow

**Copy Data Activities:** Copy\_lakehouse\_Dim\_Nutrition, Copy\_lakehouse\_Fact\_Nutrition

**Datawarehouse Name:** HealthCare\_DW

**Schema Table Names:** dim\_nutrition, fact\_nutrition

**View Names**: vw\_DIM\_NUTRITION, vw\_FACT\_NUTRITION

**Data Model Name:** Nutrition

**PowerBI Report Name:** Demo

**Issues/Challenges**

1. Several SQL queries like alter, identity columns cannot be created.
2. Unable to create measures in PowerBI

**Step-by-Step Process**

**Approach1**

1. Data has been uploaded form local folder into the lakehouse HealthCare\_Data folder.

**A screenshot of a computer

Description automatically generated**

1. Converting the files from the HealthCare\_Data folder to delta tables using Load to Tables method.

A screenshot of a computer

Description automatically generated

1. Created a new HelathCare\_Pipeline to transfer data from Healthcare\_lakehouse to Healthcare\_DW by creating a new copy data activity

A screenshot of a computer

Description automatically generated

1. Performing Cleaning and Transformations like deleting null values etc in HealthCare\_DW using custom SQL Queries.

A screenshot of a computer

Description automatically generated

1. Creating Dimension and Fact Tables like DIM\_PACKAGE, FACT\_PRODUCT etc in HealthCare\_DW after making necessary transformations.

**A screenshot of a computer

Description automatically generated**

1. Creating custom views from dim and fact tables for Analysis purposes

**A screenshot of a computer

Description automatically generated**

1. Creating PowerBI stories from the views

* **Story1:**

**A screenshot of a graph

Description automatically generated**

* **Story2:**

**A screenshot of a medical analysis

Description automatically generated**

**Approach2**

1. Loading data from local folder to HealthCare\_Nutrition\_test\_Dataflow using upload file feature

**A screenshot of a computer

Description automatically generated**

1. Making transformations like Replacing values, Change columns data, Marking keys etc in HealthCare\_Nutrition\_test\_Dataflow using transform feature

**A screenshot of a computer

Description automatically generated**

1. Creating a HealthCare\_Test\_Pipeline to perform pipeline activities

**A screenshot of a computer

Description automatically generated**

1. Creating dataflow and copy data activities in HealthCare\_Test\_pipeline

Dataflow Activity – Making transformations to the Nutritions data and post the final transformed data to lakehouse staging delta table(Nutritions\_US\_Staging)

A screenshot of a computer

Description automatically generated

1. **CopyData Activities** – Copying data from Nutritions\_US\_Staging tables to Dim and Fact Tables in Dat warehouse by mapping the source and destination schemas

A screenshot of a computer

Description automatically generated

1. Running the pipeline to insert data from Local to Lakehouse staging table to data warehouse. Pipeline has been successfully ran.

A screenshot of a computer

Description automatically generated

1. Inserting data into created FACT and DIM tables after running the pipeline activity

A screenshot of a computer

Description automatically generated

1. Creating views based on above FACT and DIM tables

A screenshot of a computer

Description automatically generated

1. Using view table column names to create stories in PowerBi.

**A blue and orange chart with numbers and a pie chart

Description automatically generated**

**Summary**

The report underscores the pivotal role of data-driven decision-making in healthcare, emphasizing its impact on patient care, resource allocation, and system efficiency. Focusing on a comprehensive dataset containing vital drug, packaging, and nutrition details, this analysis presents invaluable insights for healthcare professionals and administrators. By leveraging advanced data analysis techniques, the study aims to uncover meaningful patterns and correlations, potentially revolutionizing both patient care and the broader healthcare industry.

In terms of methodology, two distinct approaches are outlined. Approach 1 centers on meticulous data upload, transformation into delta tables, pipeline establishment, and the application of cleaning and transformation procedures. Additionally, it entails the creation of dimension and fact tables for in-depth analysis. Meanwhile, Approach 2 focuses on the loading of data, implementing crucial transformations, and establishing pipelines, dataflows, and copy data activities. This approach successfully transfers and inserts data, followed by the creation of custom views to facilitate seamless reporting and storytelling in PowerBI.