**Project Overview**

The goal of this project is to analyze the growth of customers and revenue trends also extracting insights regarding customer behavior, revenue generation and Age group analysis.

**Domain : Insurance**

**Task / KPI :**

1. Tracking customer and revenue
2. Daily Revenue and customer growth rate
3. Policy changes and monthly analysis
4. Customer Segmentation
5. Trends in customer and revenue growth
6. Age group analysis: impact of age groups on business

**Key Insights which we get:**

1. **Total Customers**: The insurance company has 26,841 customers.
2. **Total Revenue**: The company generated 989.3 million in revenue.
3. **Top City**: Delhi has the highest number of customers (11,007) and revenue (401.6 million).
4. **Top Age Group**: Customers aged 31-40 are the largest group (11,455) and generate the highest revenue (356 million).
5. **Monthly Trends**: March shows the highest growth in revenue (85%) and customers (82%), but also a significant decline of 41.7% in revenue and 41.4% in customers.
6. **Sales Channel**: The offline agent channel is the primary mode, with 55.4% of customers and 55.6% of revenue.
7. **Revenue by Sales Mode**: Revenue percentages for each sales mode are similar, ranging from 12.6% to 15.6%.
8. **Popular Policy**: Policy ID “POL4321HEL” has 4,434 customers.
9. **High-Revenue Policy**: Policy ID “POL2005HEL” generated 324.3 million in revenue.

**Links :**

Power BI :

<https://app.powerbi.com/view?r=eyJrIjoiMmYzYWI2MzYtYWQxZS00OGU0LWI1NzAtNWEyYTRjYzNiZjc3IiwidCI6ImRmODY3OWNkLWE4MGUtNDVkOC05OWFjLWM4M2VkN2ZmOTVhMCJ9>

﻿

**Azure Services and Tools Required**

1. Azure Data Lake Storage (ADLS): To store raw data.
2. Azure Data Factory (ADF): For data ingestion and ETL processes.
3. Azure Databricks: For data processing and analytics.
4. Azure SQL Database: As a data warehouse for processed data.

**Final Deliverables**

1. Azure Data Lake Storage: Organized raw and processed data.
2. Azure Data Factory Pipelines: Automated data ingestion and ETL pipelines.
3. Azure Databricks Notebooks: Code for data processing and transformation.
4. Azure SQL Database: Centralized data warehouse for analytics.
5. Power BI Dashboards: Interactive reports and visualizations.

**Column Names:**

**Azure Data Engineer role :**

**There was data ingestion, transformation, and visualization using ADF , Databricks and Azure Blob Storage**

**Pipeline Components:**

**MongoDB**: Source of the JSON data.

**Azure Data Factory (ADF):** Ingest data from MongoDB and store it in ADLS Gen2.

**Azure Data Lake Storage Gen2 (ADLS Gen2):** Store the ingested JSON data.

**Azure Databricks**: Process and transform the data using PySpark.

**Power BI**: Visualize insights and KPIs derived from the transformed data.

**Step-by-Step Pipeline:**

**Step 1:**

Data Ingestion from MongoDB to ADLS Gen2 using Azure Data Factory (ADF)

Create a Linked Service in ADF for MongoDB:

Go to Azure Data Factory.

Create a linked service for MongoDB to connect to your MongoDB instance.

Create a Linked Service in ADF for ADLS Gen2:

Create a linked service for ADLS Gen2 to connect to your data lake storage.

Create a Data Pipeline in ADF:

Create a pipeline that reads data from MongoDB and writes it to ADLS Gen2.

Use a Copy Activity to copy the data from MongoDB to ADLS Gen2.

Set the source dataset as MongoDB and the sink dataset as ADLS Gen2.

Schedule the Pipeline:

Schedule the pipeline to run at intervals to ensure data is ingested regularly.

**Step 2:**

Data Processing and Transformation using Azure Databricks

**Notebook 1:** Mounting Data from ADLS Gen2

Mount Azure Blob Storage: Mount your Azure Blob Storage where your data is located.

|  |
| --- |
| Python : |
|  |

**Notebook 2:** Transformation Logic for Business KPIs

Read and Transform Data: Read the CSV data from mounted storage, perform necessary transformations, and derive business KPIs.

|  |
| --- |
| python |
|  |