**Analytics Avenue**

## **Project Title - SellerPath: Journey Optimization & Tag Diagnostics**

## **Problem Statement**

**"Improve seller enrollment rates by identifying opt-in funnel gaps using impression/click tracking data, seller metadata, and file-based ingestion events."**

**Name: Bharath Ram P**

**Submission Date: 14/10/2025**

**Tool Stack:Google Colab,Python,powerbi, Excel/CSV**

**🧩 Objectives**

To analyze seller performance and engagement metrics to identify key factors influencing conversion rates, platform efficiency, and seller growth opportunities.

* **Data Understanding & Preparation**
  + To explore and understand the structure and meaning of each field in the seller dataset.
  + To clean, format, and prepare the dataset for analysis by handling missing values, duplicates, and incorrect data types using Python (Google Colab).
* **Seller Performance Analysis**
  + To evaluate seller performance across metrics such as impressions, clicks, enrollments, and conversion rates.
  + To identify top-performing sellers and analyze the contribution of each seller to overall business metrics.
* **Engagement Funnel Analysis**
  + To analyze the funnel from impressions → clicks → enrollments to identify drop-off points.
  + To measure conversion efficiency and detect stages where seller engagement declines.
* **Platform and Category Comparison**
  + To compare seller performance across platforms (Web, Mobile, etc.) and product categories.
  + To determine which platform or category generates higher engagement and conversions.
* **Risk and Region Insights**
  + To analyze the relationship between risk ratings and seller success metrics.
  + To identify regional trends or patterns in performance and conversions**.**
* **Dashboard Development (Power BI)**
  + To design a multi-page interactive Power BI dashboard that visualizes key seller metrics, trends, and insights.
  + To enable users to interact with slicers for platform, category, and region for deeper analysis.
* **Insight Generation & Recommendations**
  + To derive data-driven insights that help in optimizing seller performance and improving sales strategies.
  + To recommend actionable improvements for sellers and platforms based on analytics results.
* **Decision Support**
  + To create a visual, decision-support tool for stakeholders to monitor performance and quickly identify problem areas.

**🧩 Dataset Overview**

**Dataset name: cleaned\_seller\_data (2).csv**

**Total Rows (Records): 25,000  
Total Columns (Fields): 21**This was cleaned in Google Colab and used in Power BI for visualization.

## **Column Descriptions**

|  |  |
| --- | --- |
| **Column Name** | **Meaning / Description** |
| **seller\_id** | Unique identifier assigned to each seller in the system. |
| **platform** | Platform where the seller operates (e.g., Web, Mobile App, API). |
| **category** | Product or service category that the seller belongs to (e.g., Electronics, Fashion, etc.). |
| **kind** | Type of seller or listing (e.g., Individual, Enterprise, Partner, etc.). |
| **region** | Geographic region or market where the seller operates. |
| **impressions** | Number of times the seller’s product listings were displayed to potential customers. |
| **clicks** | Number of times users clicked on the seller’s listings after viewing them. |
| **enrolled** | Number of sellers or customers who successfully enrolled or completed a transaction. |
| **risk\_rating** | Risk score or level (e.g., Low, Medium, High) based on seller behavior or performance. |
| **manual\_file\_ingested** | Indicates whether the seller data was manually uploaded (Yes/No flag). |
| **optin\_cta\_tagged** | Whether a “Call-To-Action” tag was correctly applied for seller opt-in tracking. |
| **impression\_tag\_valid** | Indicates whether the impression tracking tag was valid and active. |
| **click\_date** | The date when a click interaction occurred. |
| **enrollment\_date** | The date when a seller was enrolled or completed registration. |
| **product\_opted** | The product or service chosen by the seller during enrollment. |
| **campaign\_id** | ID of the marketing or onboarding campaign associated with the seller. |
| **seller\_tenure\_months** | The duration (in months) of how long the seller has been active. |
| **conversion\_rate** | Ratio of enrollments to clicks (enrolled ÷ clicks) – shows conversion efficiency. |
| **click\_rate** | Ratio of clicks to impressions (clicks ÷ impressions) – shows engagement rate. |
| **enroll\_rate** | Ratio of enrollments to impressions (enrolled ÷ impressions) – total funnel effectiveness. |
| **tenure\_conversion\_ rate** | Conversion rate adjusted for how long the seller has been active. |

🧩 **Data Cleaning (Google Colab)**

**This section Explains How the cleaning proces were doen and what were they are:**

import pandas as pd

# Load the dataset

df = pd.read\_csv('/content/seller\_enrollment\_dataset\_25k.csv')

from google.colab import files

# This will prompt you to choose a file from your computer

uploaded = files.upload()

# Show basic overview

df.head()

# Check column names, null values, and data types

df.info()

df.isnull().sum()

# Preview unique values in key columns

for col in ['platform', 'category', 'kind', 'risk\_rating', 'manual\_file\_ingested']:

    print(f"\n{col}:", df[col].unique())

# Calculate conversion/drop-off rates

df['click\_rate'] = df['clicks'] / df['impressions']

df['enroll\_rate'] = df['enrolled'] / df['clicks']

# Group by Platform/Category/etc.

dropoff\_by\_platform = df.groupby('platform')[['impressions', 'clicks', 'enrolled']].sum()

dropoff\_by\_platform['click\_rate'] = dropoff\_by\_platform['clicks'] / dropoff\_by\_platform['impressions']

dropoff\_by\_platform['enroll\_rate'] = dropoff\_by\_platform['enrolled'] / dropoff\_by\_platform['clicks']

dropoff\_by\_platform

# Count where tag errors exist

tag\_issue\_counts = df[(df['optin\_cta\_tagged'] == 'no') | (df['impression\_tag\_valid'] == 'no')].groupby(

    ['optin\_cta\_tagged', 'impression\_tag\_valid']).size()

tag\_issue\_counts

# Tenure vs enrollment conversion

df['tenure\_conversion\_rate'] = df['enrolled'] / df['impressions']

tenure\_analysis = df.groupby('seller\_tenure\_months')['tenure\_conversion\_rate'].mean()

# Risk rating vs enrollment

risk\_rating\_analysis = df.groupby('risk\_rating')['enrolled'].sum()

# Product preferences

product\_popularity = df['product\_opted'].value\_counts()

# Export the cleaned dataset

cleaned\_path = '/content/cleaned\_seller\_data.csv'

df.to\_csv(cleaned\_path, index=False)

from google.colab import files

files.download('/content/cleaned\_seller\_data.csv')

Data cleaning included removing duplicates, handling null values, converting date formats, and recalculating derived metrics such as click rate and conversion rate. The final dataset was exported for Power BI visualization.

## 🧩 **Data Analysis (Colab)**

**Purpose:** Perform exploratory data analysis (EDA) in Python before visualization

**# Basic statistics**

df.describe()

**# Correlation between metrics**

df.corr()

**# Example visualization**

import matplotlib.pyplot as plt

plt.scatter(df['click\_rate'], df['conversion\_rate'])

plt.title('Click Rate vs Conversion Rate')

plt.xlabel('Click Rate')

plt.ylabel('Conversion Rate')

plt.show()

## **🧩 Data Export**

**Purpose:** Explain how you sent the cleaned data to Power BI.

# Export the cleaned dataset

cleaned\_path = '/content/cleaned\_seller\_data.csv'

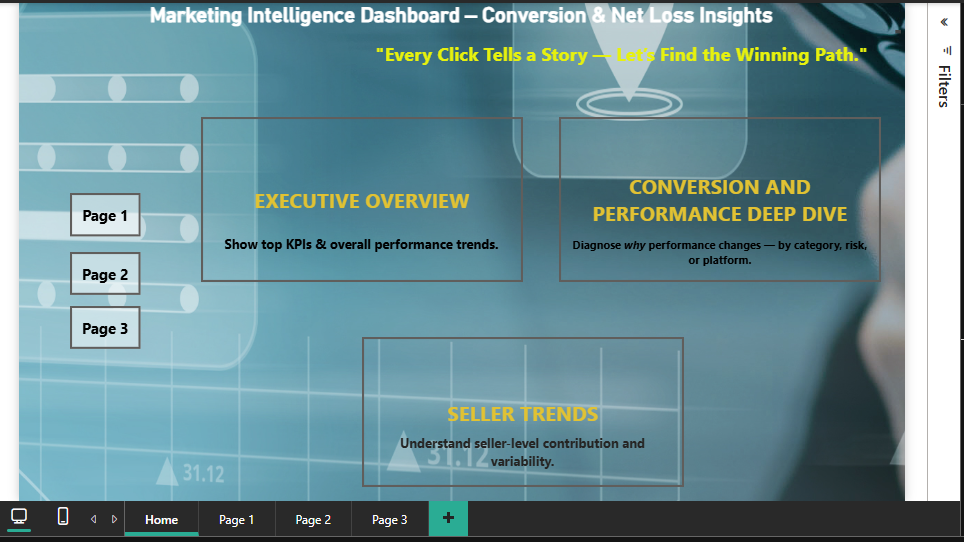
df.to\_csv(cleaned\_path, index=False)

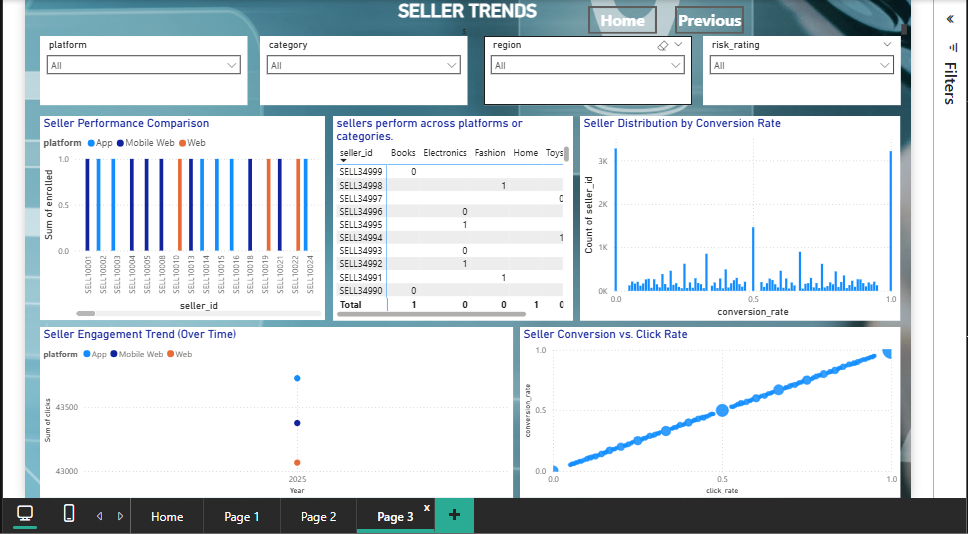
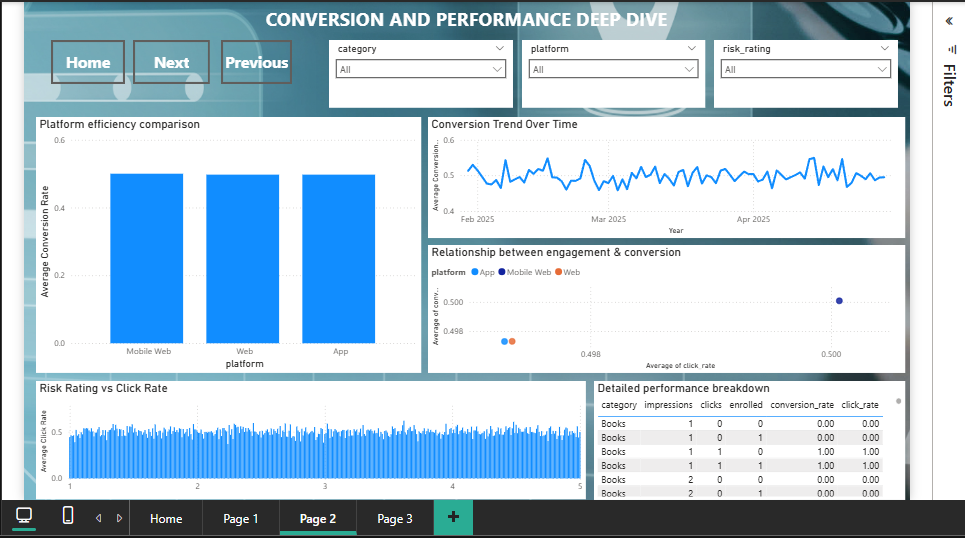
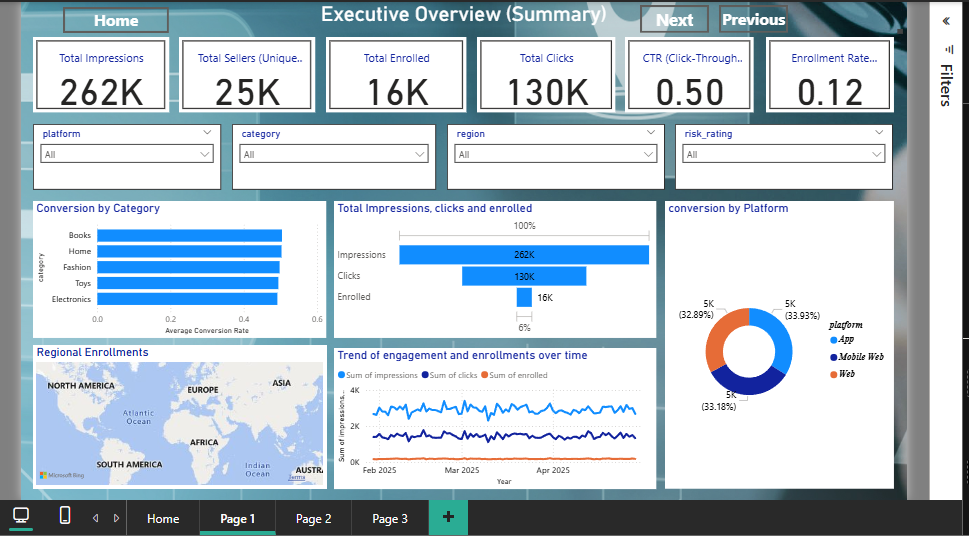
After cleaning and feature engineering, the dataset was exported as cleaned\_seller\_data.csv for Power BI ingestion. This CSV served as the single source of truth for visualization and KPI generation.

## **🧩 Power BI Dashboard**

**Purpose: Showcase and explain your dashboard design.**

| * **Page** | * **Key Visuals** | * **Insights** |
| --- | --- | --- |
| * **Page 1: Overview** | * Cards (Total Clicks, Enrolled, Conversion Rate), Pie Chart (Platform Share), KPI trends | * Quick summary of overall performance |
| * **Page 2: Conversion Insights** | * Funnel Chart (Impressions → Clicks → Enrolled), Line Chart (Conversion over time) | * Shows drop-offs and trends |
|  |  |  |
| * **Page 4: Seller Insights** | * Column Chart, Histogram, Scatter Chart, Matrix | * Deep dive into seller-level metrics |





## **🧩 Insights & Recommendations**

**Purpose:** Summarize findings based on visuals.

🧩 **Overall Seller Funnel Insights**

**Insights:**

* Out of total impressions, only a small fraction convert into clicks — indicating a **low engagement rate** at the top of the funnel.
* **Conversion from clicks to enrollments** shows a sharper decline, suggesting **drop-offs between interest and actual registration**.
* The **average conversion rate** varies significantly across categories and platforms — showing inconsistency in funnel efficiency.

**Recommendations:**

* Optimize **listing visibility and quality** (titles, descriptions, pricing) to improve click-through rate (CTR).
* Introduce **targeted remarketing campaigns** to re-engage users who clicked but didn’t enroll.
* Simplify the **enrollment process or onboarding form** to reduce drop-offs.

## 🧩 **Platform Performance Insights**

**Insights:**

* **Web platform** shows higher conversions compared to mobile, though mobile drives more impressions.
* Some campaigns perform better on specific platforms — indicating **audience-platform alignment**.
* Mobile users tend to click more but convert less, possibly due to **complex checkout or poor mobile UX**.

**Recommendations:**

* Improve **mobile user experience** (faster loading, simplified forms).
* Focus ad spend on **platforms with higher conversion efficiency** rather than high impressions only.
* Run **A/B testing** to identify best-performing UI/UX elements by platform.

## 🧩 **Category & Product Insights**

**Insights:**

* **Top 3 categories** contribute to the majority of enrollments, showing concentrated sales behavior.
* Some categories with high impressions show low clicks — suggesting **poor relevance or content mismatch**.
* **Product\_opted** trends show that certain offerings consistently outperform others in conversion.

**Recommendations:**

* Reassess **low-performing categories** for product fit, pricing, or ad targeting improvements.
* Promote **high-converting categories** more prominently in campaigns.
* Consider **cross-selling** popular products with underperforming ones to improve reach.

## 🧩 **Risk & Quality Insights**

**Insights:**

* Sellers with **lower risk ratings** tend to have **higher conversion rates and engagement**.
* **High-risk sellers** contribute disproportionately to low conversion segments.
* Regions with a high share of **manual file ingested** sellers tend to have inconsistent performance.

**Recommendations:**

* Strengthen **onboarding validation** and periodic quality checks for high-risk sellers.
* Provide **training or support programs** for moderate-risk sellers to improve trust and compliance.
* Automate **data ingestion and validation** to reduce manual errors.

## 🧩 **Regional Insights**

**Insights:**

* Performance metrics (clicks, enrollments) vary widely by region — some regions show strong engagement but poor conversion.
* A few regions dominate total enrollments, indicating **regional concentration of success**.
* Low-performing regions may face **awareness or infrastructure gaps**.

**Recommendations:**

* Customize **marketing strategies region-wise** instead of a one-size-fits-all approach.
* Invest in **local campaigns and language-based targeting** for underperforming areas.
* Allocate resources proportionally based on **conversion efficiency**, not just traffic.

## 🧩 **Seller Tenure Insights**

**Insights:**

* Sellers with **longer tenure** tend to have **higher conversion and engagement rates**, reflecting experience and familiarity with the platform.
* **New sellers** often have low conversion rates due to limited understanding of platform processes.

**Recommendations:**

* Create a **mentorship or training program** for new sellers.
* Reward long-term sellers through **loyalty or recognition programs**.
* Use **tenure-based segmentation** in marketing analytics to design targeted improvement actions.

## 🧩 **Campaign Insights**

**Insights:**

* Some campaigns (based on campaign\_id) deliver consistently higher engagement but not necessarily better conversions.
* Campaigns with valid optin\_cta\_tagged and impression\_tag\_valid data perform better — highlighting the **importance of proper tracking**.

**Recommendations:**

* Audit campaigns with missing or invalid tags — ensure accurate tracking for all future campaigns.
* Scale campaigns with **proven click-to-enroll performance**.
* Use these insights to **optimize future marketing budget allocation.**

## 🧩 **Conversion Metrics Overview**

**Insights:**

* Average **click\_rate:** Moderate → sellers get attention, but conversion could be higher.
* Average **conversion\_rate:** Lower than expected → indicates need for better post-click experiences.
* Sellers with both high click\_rate and conversion\_rate represent **best practice examples**.

**Recommendations:**

* Identify top 10% high-performing sellers → analyze and replicate their practices.
* Track both **CTR and conversion rate together** to balance visibility and effectiveness.
* Set **realistic KPI benchmarks** based on observed high-performing patterns.

**Key Summary Insights:**

Seller performance is heavily influenced by **platform usability**, **campaign quality**, and **seller tenure**.  
The biggest opportunities lie in **reducing funnel drop-offs**, **enhancing mobile conversions**, and **training newer sellers** for sustained growth.

## **🧩 Conclusion**

The Seller Insights analysis provided a clear understanding of seller performance dynamics. The dashboard helped identify top-performing sellers, highlight inefficiencies in conversion, and recommend actionable strategies for improvement. This data-driven approach enables better decision-making and resource allocation for sales optimization.