

DATA ANALYSIS AND THE ELITE GLOBAL SALES OUTLETS

"REVOLUTIONIZING INTERNATIONAL COMMERCE WITH CUTTING-EDGE"

PROJECT PURPOSE

❖ **The objective of this project is to perform a comprehensive data analysis on the ELITE GLOBAL AI SALES OUTLETS dataset.**

The dataset contains information on sales across various regions, markets, and stores, categorized by trade and business lines over several fiscal periods.

The analysis aims to identify key trends, patterns, and insights that can inform business strategies and decision-making at Elite Global AI.

INTRODUCTION

❖ **The ELITE GLOBAL AI SALES OUTLETS dataset comprises data across various regions, markets, stores, and business lines. The data underwent a thorough cleaning process to ensure accuracy and completeness.**

The subsequent analysis focuses on revenue and unit sales metrics, utilizing descriptive statistics, pivot table summaries, regression and Power BI for data cleaning and dashboard visualization.

METHODOLOGY

- **Data Cleaning and Preparation:** Using Power BI Query Editor, the dataset underwent rigorous cleaning to remove null values, ensure that data quality, and achieve 100% completeness. Column quality, profile, and distribution checks were performed.
- **Descriptive Statistical Analysis:** Descriptive statistics were computed using Excel Data Analysis tools to summarize key metrics such as mean, median, mode, standard deviation, skewness, and kurtosis for revenue and units sold. This provided a detailed overview of the dataset's characteristics.
- **Exploratory Data Analysis (EDA):** Pivot tables were created to summarize total revenue by year, month, line of business, and region. The analysis revealed significant insights into revenue distribution and sales performance across different segments.
- **Regression Analysis:** Regression analysis was conducted using Python to examine the relationship between units sold and revenue. Key statistics such as slope, intercept, R-squared, and standard error were computed, providing insights into the linear relationship between these variables.
- **Statsmodels Library in Python:**
 - **Purpose:** Detailed Statistical Analysis
 - **Details:** Used to perform OLS (Ordinary Least Squares) regression and extract detailed statistics such as R-squared, Adjusted R-squared, Standard Error, and more.

TOOLS USED

➤ **Power BI Query Editor:**

- **Purpose:** Data Cleaning & Visualization
- **Details:** Used for cleansing the dataset by removing rows and columns with null values. Column quality, profile, and distribution checks were performed to ensure 100% data integrity.

➤ **Excel Data Analysis Tools:**

- **Purpose:** Descriptive Analysis and Visualization
- **Details:** Utilized to calculate descriptive statistics such as mean, median, mode, standard deviation, skewness, kurtosis, and other key metrics for revenue and unit sales. Pivot tables were also created to summarize total revenue by year, month, line of business, and region.

➤ **Python (with libraries such as pandas, numpy, matplotlib, and scikit-learn):**

Purpose: Regression Analysis and Visualization

- **Details:**
 - **pandas:** Used for data manipulation and analysis.
 - **numpy:** Utilized for numerical operations and calculations.
 - **matplotlib:** Used for plotting and visualizing data.

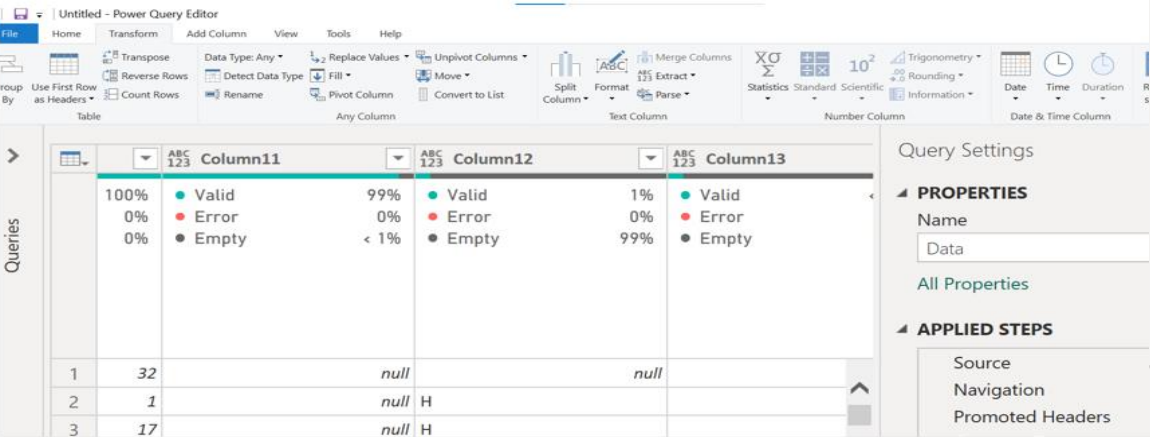
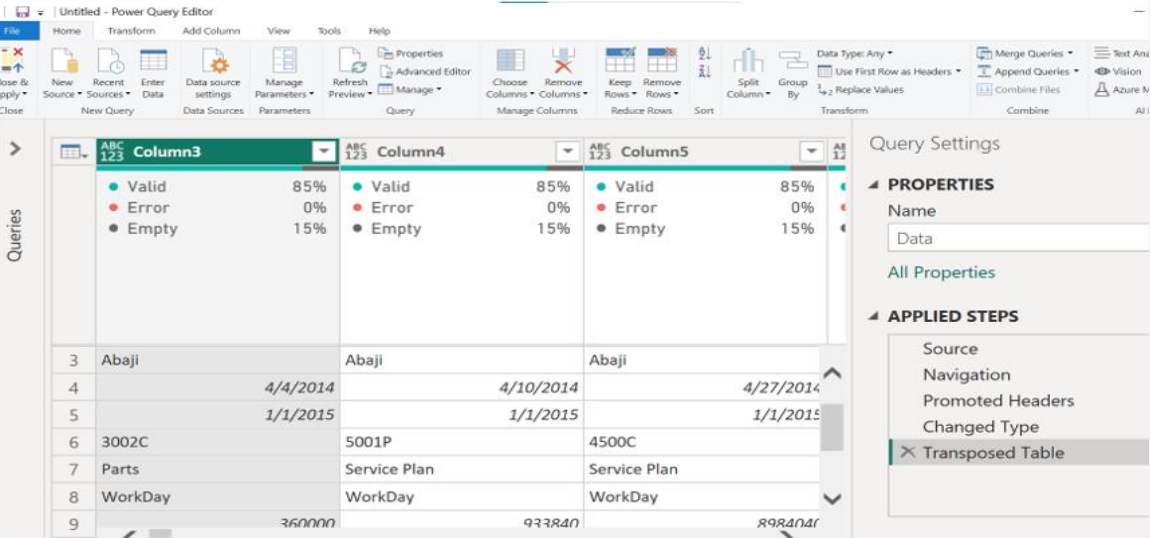
➤ **scikit-learn:** Applied for performing regression analysis to examine the relationship between units sold and revenue.

➤ **Stats models Library in Python:**

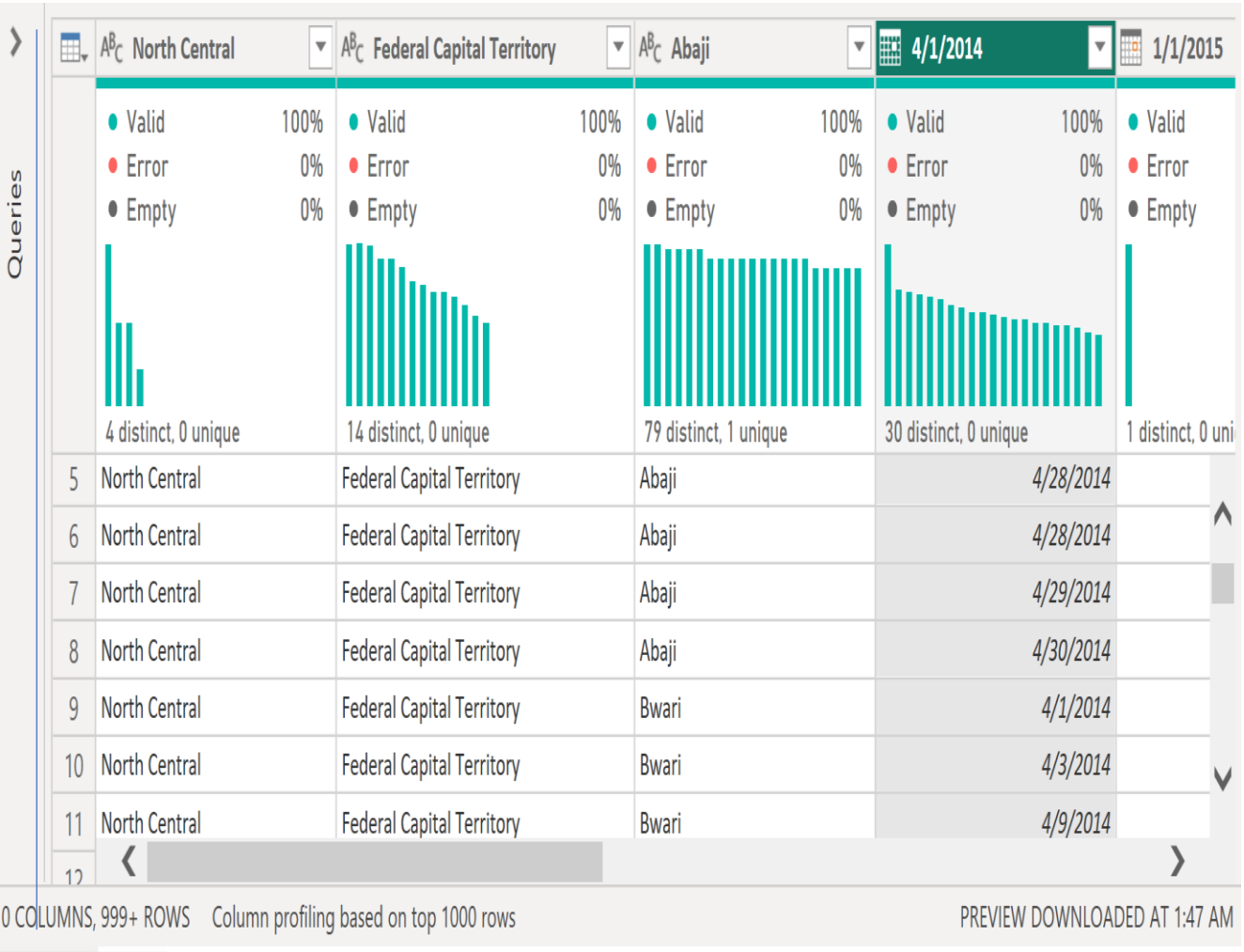
- **Purpose:** Detailed Statistical Analysis
- **Details:** Used to perform OLS (Ordinary Least Squares) regression and extract detailed statistics such as R-squared, Adjusted R-squared, Standard Error, and more.

PART OF THE DATA CLEANSING PROCESS

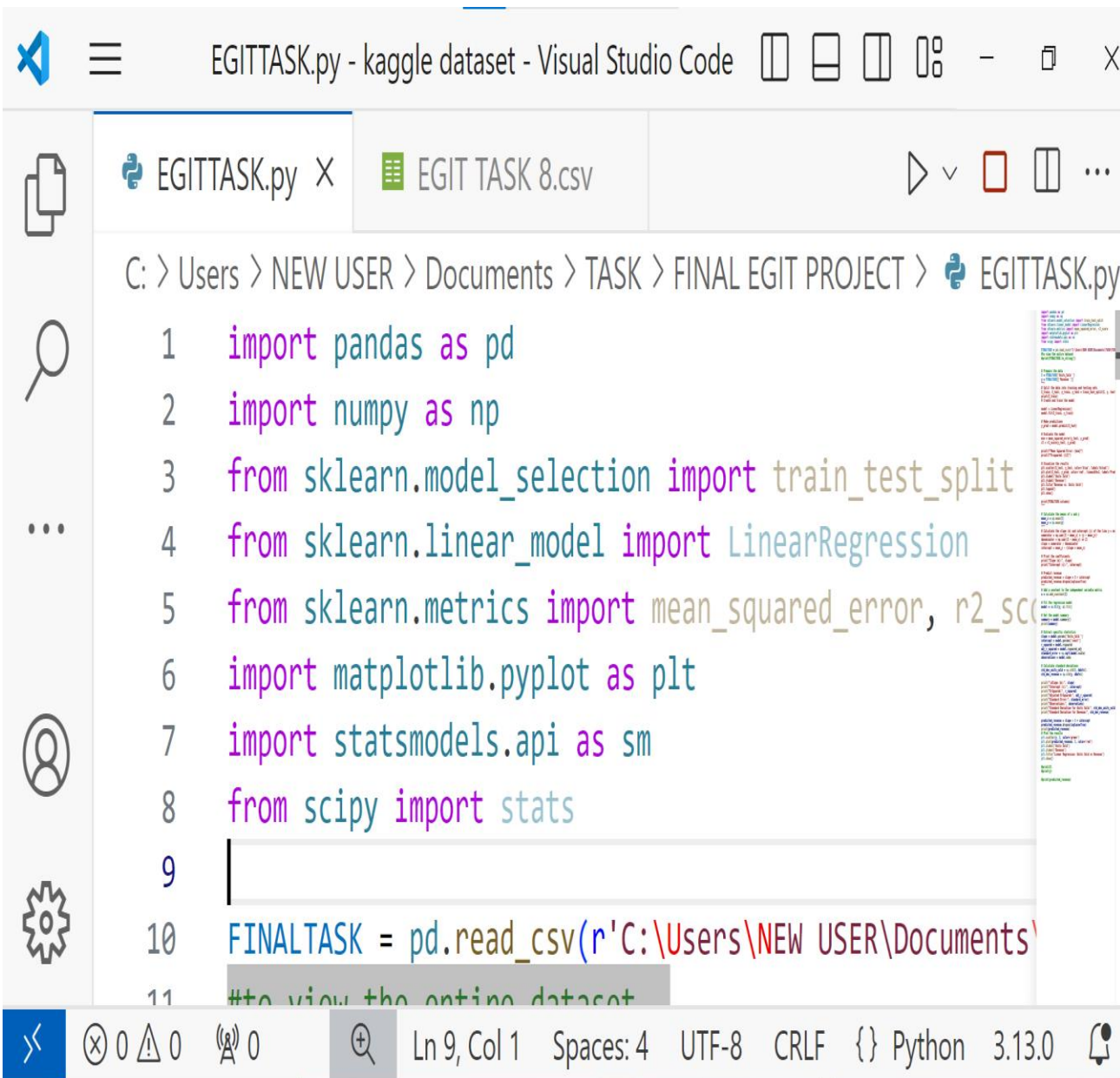
UNCLEAN DATASET, READY FOR CLEANSING PROCESS



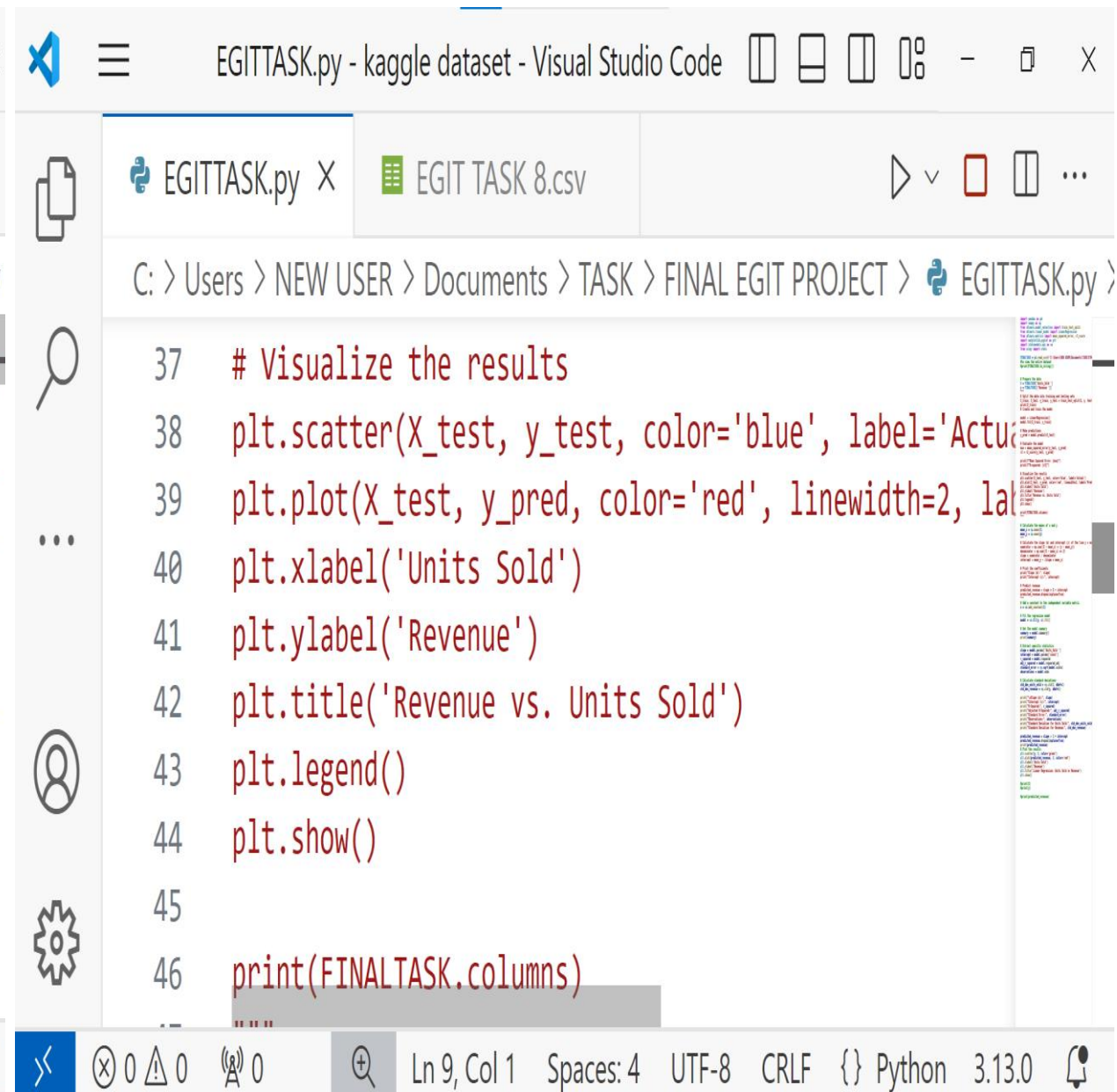
CLEANED DATASET READY FOR USAGE



CODING PROCESS USING PYTHON



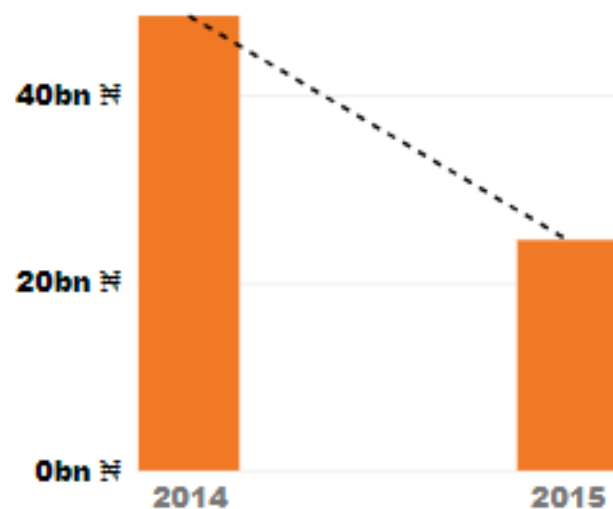
```
EGITTASK.py X EGIT TASK 8.csv
C:\Users\NEW USER\Documents\TASK\FINAL EGIT PROJECT\EGITTASK.py
1 import pandas as pd
2 import numpy as np
3 from sklearn.model_selection import train_test_split
4 from sklearn.linear_model import LinearRegression
5 from sklearn.metrics import mean_squared_error, r2_score
6 import matplotlib.pyplot as plt
7 import statsmodels.api as sm
8 from scipy import stats
9
10 FINALTASK = pd.read_csv(r'C:\Users\NEW USER\Documents\
11 #to view the entire dataset
```



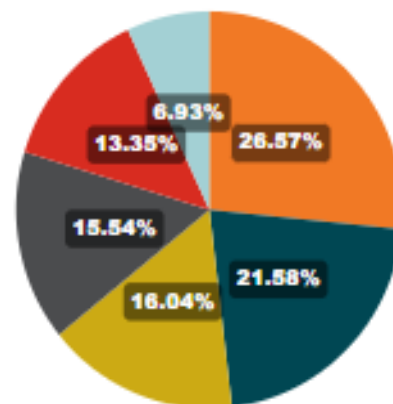
```
EGITTASK.py X EGIT TASK 8.csv
C:\Users\NEW USER\Documents\TASK\FINAL EGIT PROJECT\EGITTASK.py
37 # Visualize the results
38 plt.scatter(X_test, y_test, color='blue', label='Actual')
39 plt.plot(X_test, y_pred, color='red', linewidth=2, label='Predicted')
40 plt.xlabel('Units Sold')
41 plt.ylabel('Revenue')
42 plt.title('Revenue vs. Units Sold')
43 plt.legend()
44 plt.show()
45
46 print(FINALTASK.columns)
```

ELITE GLOBAL AL SALES OUTLETS DASHBOARD

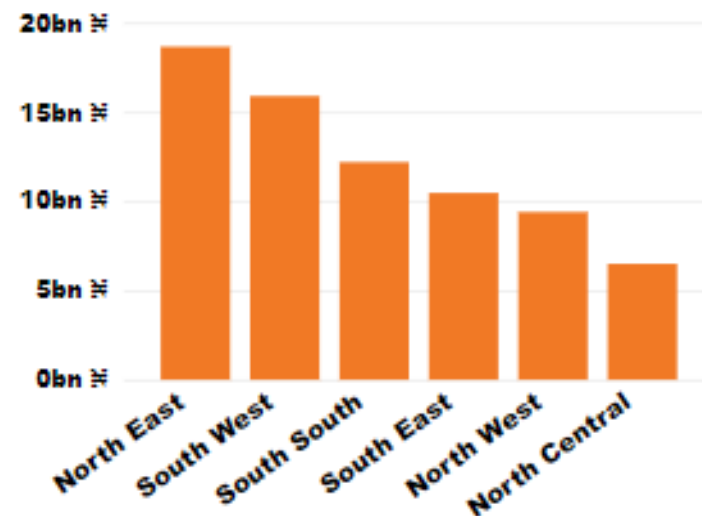
ANNUAL TOTAL REVENUE



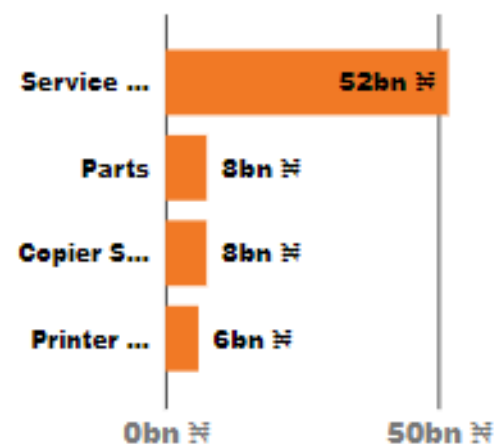
% OF UNIT SOLD BY REGION



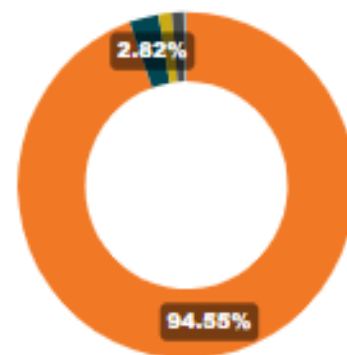
TOTAL REVENUE BY REGION



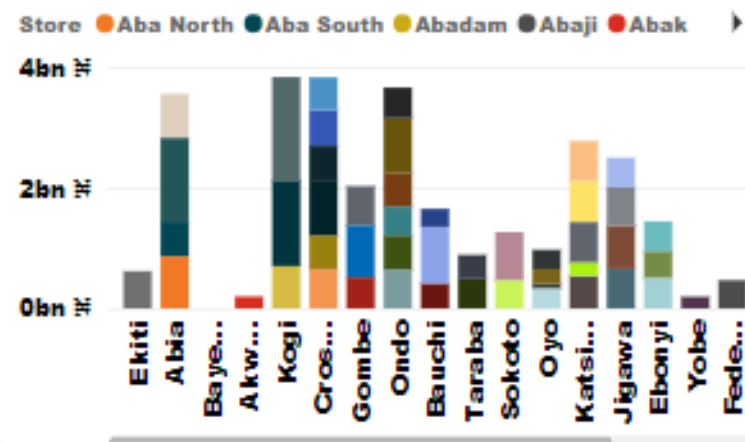
TOTAL REVENUE BY LINE OF BUSINESS



% OF UNIT SOLD BY REGION



TOTAL REVENUE GENERATED BY EACH OF THE MARKET(STORE)



TOTAL REVENUE & TOTAL UNIT SOLD BY REGION

Region	Sum of Revenue	Sum of Units Sold
North East	18,641,963,520 ₦	208,983
South West	15,877,529,040 ₦	169,730
South South	12,179,415,120 ₦	126,189
South East	10,457,403,240 ₦	122,240
North West	9,386,346,720 ₦	105,014
North Central	6,489,332,640 ₦	54,522
Total	73,031,990,280 ₦	786,678

TOTAL REVENUE & TOTAL UNIT SOLD BY LINE OF BUSINESS

Line Of Business	Sum of Revenue	Sum of Units Sold
Copier Sale	7,544,702,760 ₦	22,291
Parts	7,564,215,600 ₦	493,826
Printer Sale	6,056,589,960 ₦	12,795
Service Plan	51,866,481,960 ₦	257,766
Total	73,031,990,280 ₦	786,678

THE ELITE GLOBAL AL OUTSALES OUTLET PERFORMANCE SUMMARY TABLE

Years	Trade Date	MONTHLY INCOME
2014	Jan	3,922,133,760
	Feb	4,090,898,880
	Mar	4,157,878,920
	Apr	4,075,070,400
	May	4,213,543,680
	Jun	4,044,528,600
	Jul	3,976,032,960
	Aug	3,958,927,800
	Sep	3,876,659,640
	Oct	4,133,453,760
	Nov	4,049,343,480
	Dec	3,966,136,320
2014 Total		48,464,608,200
2015	Jan	4,049,115,360
	Feb	4,017,994,560
	Mar	4,300,901,520
	Apr	4,045,247,520
	May	4,135,105,320
	Jun	4,019,017,800
2015 Total		24,567,382,080
Grand Total		73,031,990,280
TOTAL UNIT SOLD BY LINE OF BUSINESS		
LINE OF	Sum of Units Sold	TOTAL REVENUE
Copier Sale	22,291.0	7,544,702,760
Parts	493,826.0	7,564,215,600
Printer Sale	12,795.0	6,056,589,960
Service Plai	257,766.0	51,866,481,960
Grand Tota	786,678.0	73,031,990,280

REGION	TOTAL REVENUE	STORES	10 TEN STORES BY TOTAL REVENUE
North Central	6,489,332,640	Ankpa	1,713,355,320
North East	18,641,963,520	Ajaokuta	1,424,925,840
North West	9,386,346,720	Arochukwu	1,395,110,160
South East	10,457,403,240	Ekiti South-Wes	1,248,197,880
South South	12,179,415,120	Nembe	1,086,659,280
South West	15,877,529,040	Ekiti East	1,068,455,160
Grand Total	73,031,990,280	Ado-Odo/Ota	1,013,133,840
AVERAGE REVENUE BY REGION		Dukku	965,132,760
REGION	Average of Revenue	Bauchi	940,605,840
North Central	3,517,253	Askira/Uba	904,543,200
South South	2,503,992	Grand Total	11,760,119,280
South West	2,429,614	5 TOP MARKET BY REVENUE	
North West	2,248,227	STORES	TOTAL REVENUE
North East	2,182,644	Ekiti	5,574,169,800
South East	2,161,067	Abia	5,003,724,120
Grand Total	2,371,246	Bayelsa	4,239,833,280
TOTAL UNIT SOLD BY REGION		Akwa Ibom	4,101,650,880
REGION	Sum of Units Sold	Kogi	3,841,569,720
North Central	54,522	Grand Total	22,760,947,800
North East	208,983	BOTTOW 5 STORES BY UNIT SOLD	
North West	105,014	STORES	Sum of Units Sold
South East	122,240	Boki	2
South South	126,189	Kwali	4
South West	169,730	Chibok	42
Grand Total	786,678	Isiala Ngwa Sout	54
		Akinyele	1,448
		Grand Total	1,550

INSIGHT FROM THE ABOVE ELITE GLOBAL AL SALES OUTLETS VISUALIZATION

- **Top Selling Stores by Total Revenue**

- Ankpa: The top-selling store, generating 1,713,355,320.
- **Ajaokuta:** The second top store, with 1,424,925,840.
- Arochukwu, Ekiti South-West, and Nembe follow, each generating over 1 billion in revenue.

- **Top Selling Business Line**

- Service Plan: Dominates with 51,866,481,960 in revenue.
- Parts and Copier Sale: Both contribute significantly, with over 7 billion each.
- Printer Sale: Adds another substantial amount of 6,056,589,960.

- **Region with Highest Sales**

- North East: Leads in total revenue, achieving 18,641,963,520.
- South West: Comes next with 15,877,529,040.

- **Region with Lowest Sales**

- North Central: Registers the lowest total revenue at 6,489,332,640.

- **Average Revenue by Region**

- **North Central:** Has the highest average revenue per store at **3,517,253**.
- **South East:** Has the lowest average revenue per store at **2,161,067**.

- **Total Units Sold by Region**

- **North East:** Tops in units sold with **208,983** units.
- **North Central:** Reports the lowest units sold at **54,522** units.

- **Bottom 5 Stores by Unit Sold**

- **Boki:** Sells the fewest units, only **2**.
- Other stores with low sales include **Kwali (4)**, **Chibok (42)**, **Isiala Ngwa South (54)**, and **Akinyele (1,448)**.

✓ **These insights highlight the performance disparities across regions, the dominance of certain business lines, and the specific stores that excel or struggle**

SIMPLE LINEAR REGRESSION MODEL

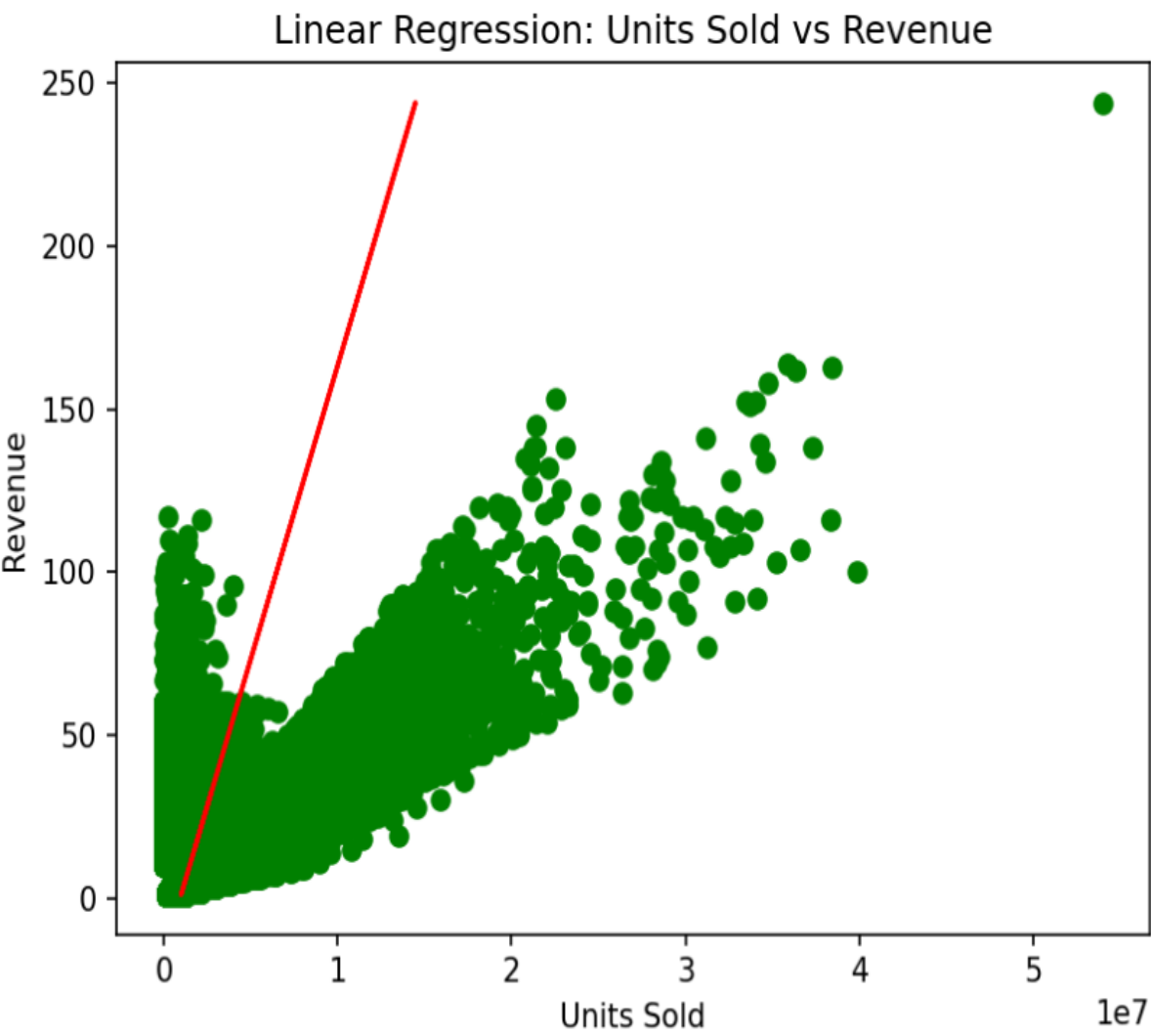
$$Y = mX + c$$

Where:

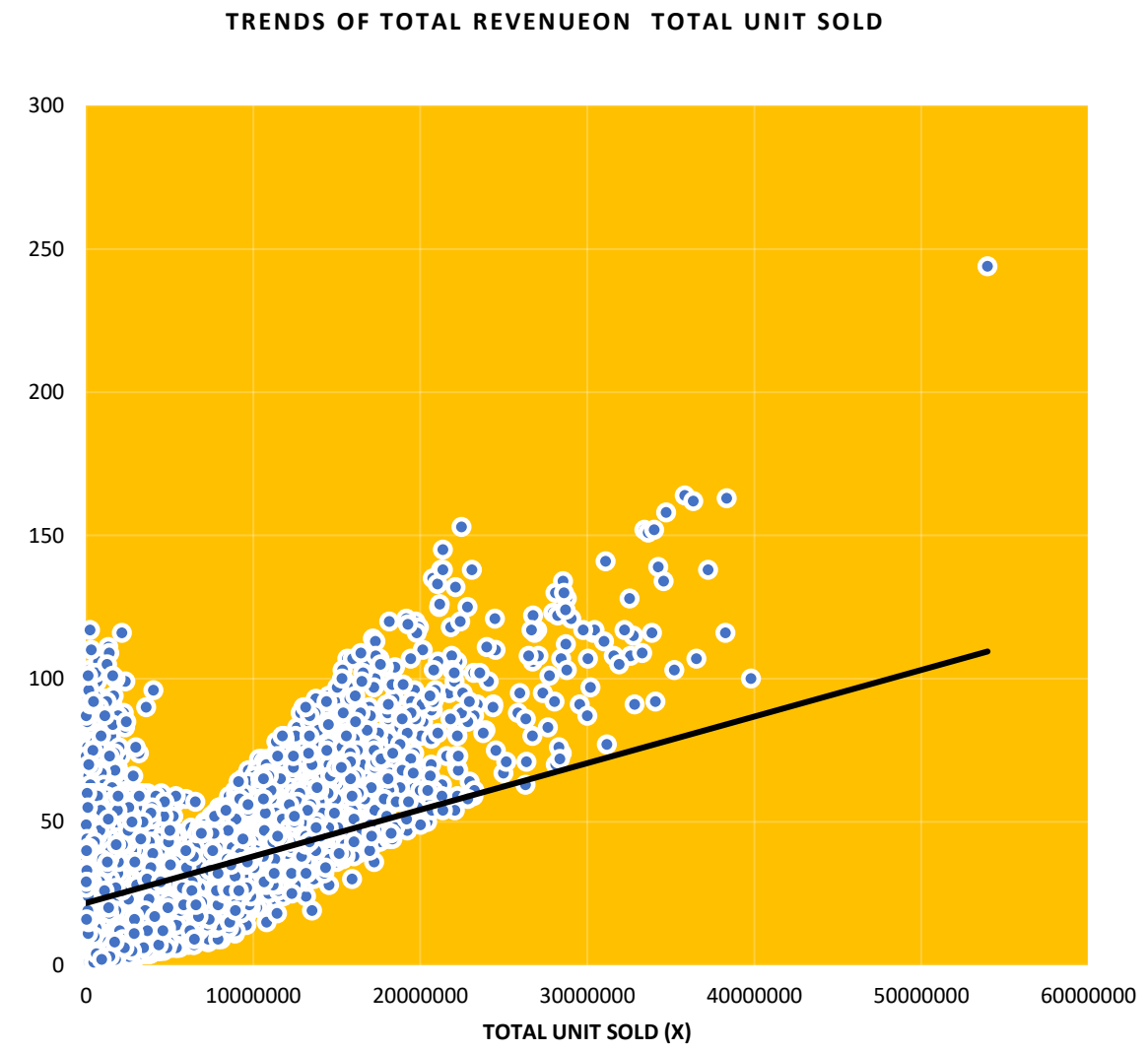
- ☐ REVENUE (Y) (dependent variable)
- ☐ UNIT SOLD (X) (independent variable)
- ☐ Slope(M) (represents the rate at which Y changes for each unit change in X
- ☐ Intercept (Total Revenue when the number of units sold is zero)

COMPARISON AND TRENDS ON PREDICTED MODEL

SIMPLE LINEAR REGRESSION CHART WITH PREDICTED VARIABLES
USING PYTHON



SIMPLE LINEAR REGRESSION LINE USING MICROSOFT EXCEL
ACTUAL DATA



REGRESSION RESULTS

➤ **Simple linear regression using predicted value (python)**

- ✓ Slope: 55,409.59
- ✓ Intercept: 955,955.92
- ✓ R-Squared: 0.0901
- ✓ Standard Error: 3,491,300.36
- ✓ Adjusted R-Square: 0.0901302074
- ✓ Standard error: 3491300.3628525054
- ✓ Observations: 30799
- ✓ Standard deviation (unit sold): 19.830825985594984
- ✓ Standard deviation (revenue): 3.660080e+06

➤ **Simple linear regression using actual values (Microsoft Excel)**

- Slope: 55409.59
- ✓ Intercept: 955955.92
- ✓ R-Squared: 0.0901
- ✓ Standard Error: 3491300
- ✓ Adjusted R-Squared: 0.0901
- ✓ Standard Deviation (unit sold) : 19.8305
- ✓ Standard Deviation (Revenue): 3660020
- ✓ Observation: 30799

REGRESSION DATA ANALYSIS KEY INSIGHTS

- **Slope (m): 55,409.59**

This indicates that for every additional unit sold, revenue increases by approximately **55,409.59**.

- **Intercept (c): 955,955.92**

When the number of units sold is zero, the total revenue is estimated to be **955,955.92**.

This is the baseline revenue when no units are sold.

- **R-Squared (0.0901)**

The R-squared value of **0.0901** indicates that approximately **9%** of the variability in revenue can be explained by the number of units sold. This suggests that there are other factors affecting revenue not captured by this model.

- **Standard Error: 3,491,300.36**

This is a measure of the accuracy of the predictions made by the model. A high standard error compared to the mean revenue suggests that the model's predictions have significant variability.

- **Adjusted R-Squared: 0.0901302074**

Similar to the R-squared value, this adjusted value also indicates a low proportion of variability explained by the model, which remains consistent when adjusting for the number of predictors in the model.

- **Observations: 30,799**

This is the total number of data points used in the analysis, providing a robust sample size.

- **Standard Deviation (Unit Sold): 19.83**

This shows the spread or dispersion of the number of units sold. A higher standard deviation indicates greater variability in units sold.

- **Standard Deviation (Revenue): 3,660,080**

This shows the spread or dispersion of revenue, indicating a wide range of revenue values.

- **Multiple R (0.3002)**

This value is the correlation coefficient, indicating a moderate positive linear relationship between units sold and revenue.

- **Analysis Summary:**

While the slope and intercept provide useful insights into the relationship between units sold and revenue, the low R-squared value suggests that the model doesn't fully capture all the factors influencing revenue. This implies that other variables might need to be considered to improve the model's predictive power.

EXPLANATION OF FINDING

❖ Comparisons and Trends:

- **Revenue shows a high level of variability with significant skewness and kurtosis, indicating the presence of outliers and a non-normal distribution.**

The positive skew in unit sales suggests that most stores sell fewer units, with some stores achieving very high sales.

- **The regression analysis indicates that while there is a positive relationship between units sold and revenue, the R-squared value is low, suggesting that many other factors affect revenue.**

❖ Benchmarks:

- **The standard deviations and variances provide benchmarks for understanding the dispersion of revenue**

CONCLUSION

- **In conclusion, the project equips Elite Global AI with the necessary insights and tools to optimize its strategic planning, enhance its market position, and drive sustainable growth. By leveraging these insights, the company can make informed decisions that align with their overall strategic goals and adapt to market dynamics efficiently.**



RECOMMENDATIONS

STRATEGIC ACTIONS

1.Focus on High-Performing Regions and Stores:

- ✓ **Prioritize investment and marketing efforts in regions like North East and South West that show higher average revenue.**
- ✓ **Investigate and replicate successful strategies from top-performing markets like**
 - **Ekiti and Abia.**

2. Optimize Inventory and Sales Strategies:

- ✓ **Utilize the insights from unit sales distribution to better manage inventory, ensuring high-demand items are always in stock.**
- ✓ **Explore targeted promotions for regions and stores with lower sales to boost performance.**

3. Enhance Data-Driven Decision Making:

- ✓ **Continue using data analytics to track performance and identify emerging trends.**
- ✓ **Invest in advanced analytics tools to uncover deeper insights into factors driving sales and revenue.**

4. Training and Development:

- ✓ **Implement training programs for sales staff in lower-performing regions to improve sales techniques and customer engagement.**

5. Customer-Centric Approach:

- ✓ **Focus on customer feedback to understand needs and preferences, tailoring products and services to meet demand more effectively.**