UNIVERSITY OF EASTERN AFRICA, BARATON

P.O BOX 2500-30100 ELDORET, KENYA

SCHOOL OF BUSINESS

DEPARTMENT OF INFORMATION SYSTEM AND COMPUTING

Data Analysis Hackathon 2025: E-commerce Sales and Customer Insights

Hackathon Guide

Prepared by: Mr Jefferson Mwatati

May, 2025

Hackathon Guide

Data Analysis Hackathon 2025: E-commerce Sales and Customer Insights

This hackathon is designed to empower learners with hands-on experience in data analysis using tools like Excel, SQL, Python, and Power BI. Leaners will analyze an E-commerce Sales and Customer Insights Dataset to uncover trends, answer key business questions, and create actionable insights. The project will culminate in a GitHub repository showcasing your work, including SQL queries, Python scripts, visualizations, and a Power BI dashboard.

Hackathon Timeline

Day 1-2: Set up GitHub, download dataset, and clean data using Excel.

Day 3-4: Export data to a database, clean and analyze data using SQL.

Day 5-6: Perform exploratory data analysis (EDA) and visualization using Python.

Day 7-8: Create a Power BI dashboard and finalize the report.

Day 9: Submit your project on GitHub before the deadline.

Step-by-Step Guide

1. Set Up GitHub Repository

- Create a GitHub profile (if you don't have one).
- Create a new repository named "*E-commerce-Sales-Analysis-Hackathon"*.
- Add a *README.md* file to document your project.
- Ensure the repository includes the following files:
- SQL_Analysis_Queries.sql: All SQL queries used for data cleaning and analysis.
- Python_Analysis_Scripts.ipynb: Python scripts for EDA and visualization.
- PowerBI_Analysis_Dashboard.pbix: Power BI dashboard file (optional to upload, but include a link).
- README.md: Detailed report with screenshots, explanations, and links.

2. Download and Clean Data

- Download the dataset from Kaggle: E-commerce Sales and Customer Insights Dataset. https://www.kaggle.com/datasets/refiaozturk/e-commerce-sales/data
- Use Excel to clean the data:
    - Remove duplicates.

- Handle missing values.
- Standardize formats (e.g., dates, categories).
- Document your cleaning process in the README.md file.

3. Export Data to a Database and Analyze Using SQL

Export the cleaned dataset to a database (e.g., MySQL, PostgreSQL, or SQLite).

- Use SQL to:
- Clean the data further (if needed).
- Perform exploratory data analysis (EDA) to answer the following questions:
  - What are the total sales by region?
  - Which product category generates the highest revenue?
  - What is the average shipping fee by region?
  - How does customer age impact purchasing behavior?
  - What is the most popular product by gender?
  - What is the order fulfillment rate (delivered vs. returned)?
- Save all SQL queries in the SQL_Analysis_Queries.sql file.
- Take screenshots of your SQL results and include them in the README.md file.

4. Perform EDA and Visualization Using Python

- Use Python (Jupyter Notebook or any IDE) to:
  - Perform additional EDA to answer the project questions.
  - Create visualizations (e.g., bar charts, pie charts, scatter plots) to showcase insights. Example visualizations:
    - Sales trends over time.
    - Distribution of sales by product category.
    - Correlation between age and total price.
- Save your Python scripts in the Python_Analysis_Scripts.ipynb file.
- Take screenshots of your visualizations and include them in the README.md file.

5. Create a Power BI Dashboard

- Use Power BI to create an interactive dashboard:
  - Include key metrics like total sales, revenue by category, and shipping status.
  - Add visualizations like maps, bar charts, and pie charts.

  - Ensure the dashboard is user-friendly and visually appealing.
  - Publish your dashboard to Power BI Service and include the link in the README.md file.
- Take a screenshot of your dashboard and include it in the README.md file.

6. Document Your Report in the README.md File

Structure your README.md file as follows:

- **Introduction:** Briefly describe the dataset and the purpose of the analysis.
- **Data Cleaning Process:** Explain how you cleaned the data using Excel and SQL.
- **SQL Analysis:** Answer the project questions using SQL. Include screenshots of your queries and results.
- **Python Analysis:** Explain your EDA process and visualizations. Include screenshots of your Python scripts and visualizations.
- **Power BI Dashboard:** Describe the insights from your dashboard. Include a screenshot and link to the dashboard.
- **Conclusion:** Summarize your findings and provide actionable recommendations for the e-commerce business.

Hackathon Rules

GitHub Repository

- Create a repository and upload all required files.
- **Do not update the repository after the deadline. Any updates will result in disqualification.**

Dataset

- Use the provided dataset from Kaggle.
- Clean the data using Excel and SQL.

Tools

- Use Excel for initial data cleaning.
- Use SQL for data analysis and cleaning.
- Use Python for EDA and visualization.
- Use Power BI for dashboard creation.

Submission

- Ensure your GitHub repository includes all required files and a well-documented README.md file.
- Submit the repository link before the deadline.

Evaluation Criteria

Completeness: All required files and steps are included.

Accuracy: Correct and meaningful analysis of the dataset.

Creativity: Unique and insightful visualizations and recommendations.

Documentation: Clear and well-structured report in the README.md file.

Technical Skills: Proficiency in using Excel, SQL, Python, and Power BI.

Hackathon Questions

Use the following questions to guide your analysis.

Sales Analysis

1. What are the total sales by region?
2. Which product category generates the highest revenue?
3. What is the average shipping fee by region?

Customer Behaviour

1. How does customer age impact purchasing behavior?
2. What is the most popular product by gender?

Order Fulfillment

1. What is the order fulfillment rate (delivered vs. returned)?
2. Are there any trends in shipping status over time?

Feel free to explore other phenomenon of data insights you identify in your analysis.

Tips for Success

1. **Learn as you go:** Use online resources to familiarize yourself with the tools.
2. **Collaborate:** Share ideas and solutions with fellow participants.
3. **Be creative:** Think outside the box to uncover unique insights.
4. **Document everything:** Ensure your report is detailed and easy to understand.

Final Submission Checklist

1. GitHub repository with all required files.
2. SQL queries and results documented.
3. Python scripts and visualizations included.
4. Power BI dashboard link and screenshot.
5. Well-written README.md file with a detailed report.

After the hackathon, you can visit this link to review the expected outcomes. https://github.com/jefftizo/E-commerce-Sales-Analysis-Hackathon