

Week 2 Tasks

Data Science Intern

SALES DATA ANALYSIS DASHBOARD

Description: Analyze a sales dataset and create visual reports showing monthly sales trends, best-selling products, and customer insights.

Challenges:

- Handle missing & inconsistent data
- Group by region/category for insights
- Create 3–5 meaningful visualizations
- **Tech Stack:** Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook

SENTIMENT ANALYSIS ON TWITTER DATA

Description: Perform text analysis on tweets (can use a pre-collected dataset). Classify tweets as positive, neutral, or negative.

Challenges:

- Clean and preprocess tweet text (remove stopwords, punctuation, etc.)
- Visualize sentiment distribution
- Use a simple model like Logistic Regression or Naive Bayes
- **Tech Stack:** Python, Pandas, NLTK/TextBlob, Scikit-learn

CUSTOMER SEGMENTATION USING CLUSTERING

Description: Use K-Means or hierarchical clustering to group customers based on purchasing behavior.

Challenges:

- Scale and normalize features
- Choose optimal k using the Elbow Method
- Visualize clusters using 2D/3D plots
- **Tech Stack:** Python, Scikit-learn, Pandas, Matplotlib

*Complete any 1 Task