AitroniX

eduzah

Eduzah

Data Analysis Internship

Data Analysis Tasks

Task 1: Netflix Shows Exploration & Insights

Dataset: Netflix Movies and TV Shows

Objective: Analyze the Netflix catalog to uncover trends in content, genres, ratings, and release patterns.

Steps:

1. Load the dataset and inspect its structure.
2. Clean missing values (ratings, director, country).
3. Analyze content type distribution (Movies vs TV Shows).
4. Identify top genres, countries, and release years.
5. Visualize trends (heatmaps, bar charts, word clouds).
6. Visualize via interactive dashboard (Power BI dashboard).

Task 2: IMDB Top 1000 Movies & TV Shows Analysis

Dataset: IMDB Movies Dataset

Objective: Explore top-rated movies & shows to find rating trends, top directors, and genre patterns.

Steps:

1. Preprocess the dataset (clean genres, directors, runtime).
2. Analyze ratings vs year of release.
3. Identify top 10 directors & actors by average rating.
4. Create genre-based insights.
5. Visualize via interactive dashboard (Power BI dashboard).

Task 3: Bank Customer Churn Prediction (EDA & Feature Analysis)

Dataset: Churn Modelling

Objective: Perform exploratory analysis to identify churn factors in banking customers.

Steps:

1. Clean data & encode categorical features.
2. Analyze churn rate across geography, gender, age, balance.
3. Visualize churn correlations (heatmap, box plots).
4. Suggest key features for predictive modeling.
5. Visualize via interactive dashboard (Power BI dashboard).

Task 4: World Happiness Report Insights

Dataset: World Happiness Dataset

Objective: Explore factors contributing to happiness across countries.

Steps:

1. Clean and normalize data.
2. Compare happiness scores by region.
3. Analyze correlation between GDP, life expectancy, freedom, and happiness.
4. Build a simple linear regression to predict happiness score.
5. Visualize top 10 happiest and least happy countries.
6. Visualize via interactive dashboard (Power BI dashboard).

Task 5: Telco Customer Churn Analysis

Dataset: Telco Customer Churn

Objective: Find key churn drivers and customer behavior patterns.

Steps:

1. Handle missing values and categorical encoding.
2. Analyze churn rate by contract type, internet service, and payment method.
3. Visualize churn patterns with interactive graphs.
4. Recommend retention strategies.
5. Visualize via interactive dashboard (Power BI dashboard).

Task 6: Heart Disease Risk Factors Analysis

Dataset: Indicators of Heart Disease (2022 UPDATE)

Objective: Analyze lifestyle and demographic factors linked to heart disease.

Steps:

1. Clean data and handle imbalances.
2. Analyze risk factors (age, smoking, BMI, physical activity).
3. Build a correlation matrix and risk profile.
4. Suggest health awareness strategies.
5. Visualize via interactive dashboard (Power BI dashboard).

Task 7: Customer Personality Analysis

Dataset: Customer Personality Analysis

Objective: Analyze customer demographics, spending habits, and marketing interactions to uncover key business insights.

Steps:

1. Clean and preprocess data.
2. Analyze demographics (age, education, marital status, household size).
3. Explore spending patterns across product categories and marketing campaigns.
4. Suggest actionable marketing and customer retention strategies based on findings.
5. Visualize via interactive dashboard (Power BI dashboard).

Task 8: Loan Approval Factors Analysis

Dataset: Loan Approval Classification Data

Objective: Analyze demographic and financial factors affecting loan approval outcomes.

Steps:

1. Clean data and handle missing values.
2. Analyze approval rates by demographics and income levels.
3. Build a correlation matrix for numeric variables.
4. Suggest strategies to improve loan approval processes and customer guidance.
5. Visualize via interactive dashboard (Power BI dashboard).

Task 9: Student Depression Risk Factor Analysis

Dataset: Student Depression Dataset

Objective: Analyze demographic, academic, and lifestyle factors influencing depression levels among students.

Steps:

1. Clean data and handle missing values.
2. Analyze risk factors (e.g., gender, academic pressure, sleep patterns). Examine differences in depression incidence across demographic groups and stressors.
3. Build a correlation matrix for numeric variables.
4. Suggest student mental health awareness strategies.
5. Visualize via interactive dashboard (Power BI dashboard).

Task 10: Video Game Sales Insights Analysis

Dataset: Video Games Sales Dataset

Objective: Analyze video game sales trends across regions, platforms, genres, publishers, and over time to uncover actionable market insights.

Steps:

1. Clean data and handle missing values.
2. Analyze sales trends by key categories.
3. Build a correlation or association overview of numeric variables.
4. Suggest strategic recommendations for stakeholders.
5. Visualize via interactive dashboard (Power BI dashboard).