Data Inspection, Analysis, and Exploratory Data Analysis (EDA)

Overview

This project demonstrates expertise in data inspection, analysis, and exploratory data analysis (EDA) using Python and Tableau. The skills applied include data wrangling, cleaning, statistical analysis, and visualisation to uncover insights from a TikTok dataset.

Part 1: Inspect and Analyze Data

Skills Applied:

- Data Handling with `pandas` and `numpy`
- Data Quality Assessment
- Statistical Summary Computation

Understanding the Situation

To analyze the TikTok dataset effectively, the following steps were taken:

- Examined dataset structure and key variables.
- Identified relevant attributes for analysis.
- Assessed data quality and addressed inconsistencies.
- Established a foundation for exploratory data analysis (EDA).

Understanding the Data

1. DataFrame Construction

- Created a structured DataFrame using `pandas` and `numpy`.
- Loaded the dataset from a CSV file for analysis.

2. Initial Data Inspection

- Displayed the first few rows with `df.head()` to get an overview.
- Used `df.info()` to understand the dataset structure and data types.
- Checked for missing values and inconsistencies.

- Calculated unique value counts using `df.nunique()` to identify categorical variables.

3. Summary Statistics

- Generated statistical summaries using `df.describe()`.
- Examined distributions, central tendencies, and variations.
- Detected potential outliers.

Understanding the Variables

- Differentiated between categorical and numerical features.
- Analyzed feature correlations to detect relationships.
- Identified redundant or irrelevant columns.
- Considered necessary transformations, such as encoding categorical variables and scaling numerical values.

Part 2: Exploratory Data Analysis (EDA)

Skills Applied:

- Data Cleaning and Transformation
- Data Visualization with `matplotlib` and `seaborn`
- Interactive Dashboarding with Tableau

Part 1: Imports, Links, and Loading

- Imported necessary Python libraries:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

- import seaborn as sns
- Successfully loaded the dataset for analysis.

Part 2: Data Exploration & Cleaning

- Identified and handled missing values using `df.isnull().sum()`.
- Detected and removed duplicate records.
- Standardised data formats and corrected inconsistencies.
- Transformed variables where necessary (e.g., converting date formats, normalising numerical values).

Part 3: Build Visualizations

Univariate Analysis

- Created histograms to explore data distributions.
- Used box plots to detect outliers.
- Generated count plots to analyse categorical variables.

Bivariate & Multivariate Analysis

- Developed scatter plots to examine relationships between numerical features.
- Constructed heatmaps to visualise correlations.
- Built bar plots to analyse trends in categorical data.
- Utilised Tableau for interactive and advanced visualisations.

Part 4: Evaluate and Share Results

- Summarised key findings and insights from the data.
- Highlighted areas requiring further investigation.
- Documented next steps include feature engineering, hypothesis testing, and model building.
- Prepared and exported cleaned data for further statistical analysis or machine learning applications