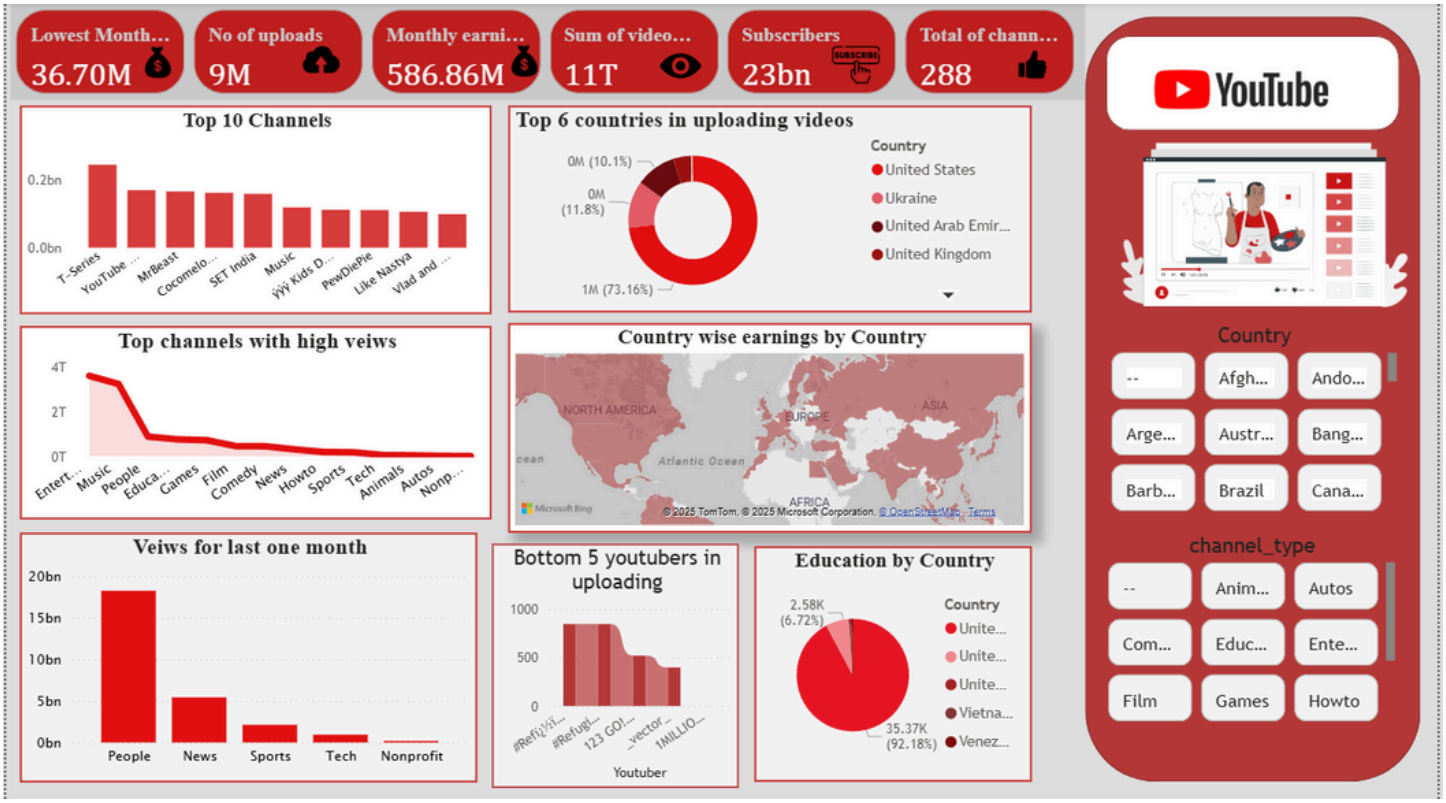
Project Summary



Key points

YouTube Data Analysis Report

Introduction

This project focuses on analyzing YouTube channel performance, subscriber base, content uploads, and revenue distribution. The data was first cleaned and preprocessed using **Python** to ensure consistency, remove duplicates, and handle missing values. After cleaning, **machine learning techniques** were applied for insights, trend identification, and categorization of channels. The results have been visualized in the dashboard above for better interpretation.

Data Cleaning and Preparation

- Removed duplicate and inconsistent channel records.
- Standardized country and channel type fields.
- Handled missing values in views, subscribers, and earnings columns.
- Converted categorical variables into suitable formats for machine learning analysis.

Machine Learning Applications

- **Clustering:** Grouped YouTube channels based on earnings, uploads, and views.

- **Classification Models:** Predicted channel performance categories (high-performing vs. low-performing).
- **Trend Analysis:** Identified growth patterns in subscriber counts and video uploads.
- **Anomaly Detection:** Detected channels with unusual earnings or subscriber spikes.

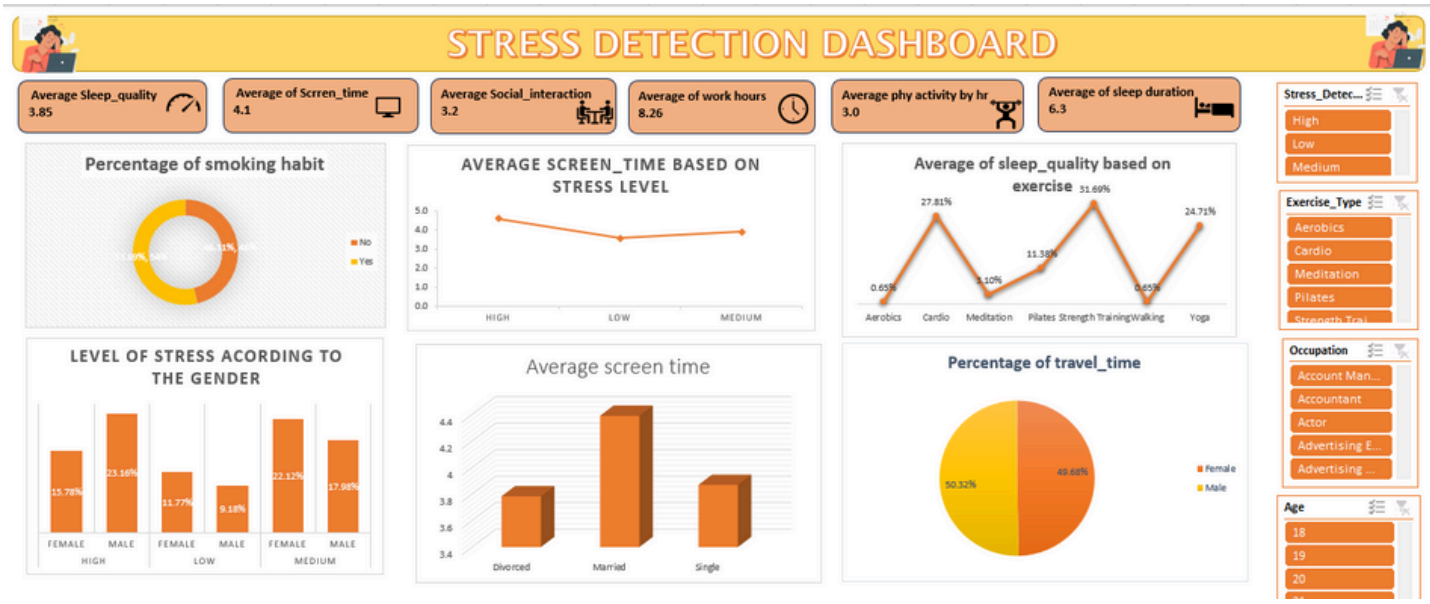
Key Insights from Dashboard

1. **General Statistics:**
   - **Lowest Monthly Earnings:** $36.70M
   - **Number of Uploads:** 9M
   - **Monthly Earnings:** $586.8M
   - **Sum of Video Views:** 11T
   - **Total Subscribers:** 23B
   - **Total Channels Analyzed:** 288
2. **Top Channels:**
   - *T-Series* leads as the most viewed and subscribed channel, followed by *YouTube*, *MrBeast*, and *Cocomelon*.
   - Entertainment, music, and kids' content dominate the top positions.
3. **Country-wise Trends:**
   - **United States** contributes the largest share in terms of video uploads (73.16%).
   - Other notable contributors include Ukraine, UAE, and the United Kingdom.
4. **High View Categories:**
   - Entertainment and Music channels generate the most views, followed by People and Education.
5. **Monthly View Distribution:**
   - *People-related content* gained the highest views in the last month (~18B views).
   - News, Sports, Tech, and Nonprofit had lower but notable viewership.
6. **Country-wise Earnings:**
   - Highest earnings are concentrated in the United States, India, and select European and Asian countries.
7. **Bottom Performing Channels:**
   - A few channels show very low upload activity, highlighting underutilization despite the platform's opportunities.
8. **Education-based Insights:**
   - Education channels are dominated by Vietnam and Venezuela, but their overall proportion is smaller compared to entertainment and music.

Conclusion

The analysis shows that YouTube is heavily dominated by entertainment, music, and people-related content, which collectively drive most of the views and earnings. The United States leads in uploads and earnings, while other countries like India and Vietnam have niche strengths. Machine learning models helped uncover trends and group channels by performance, making it easier to identify growth opportunities.

By leveraging this analysis, content creators can focus on high-engagement categories, while businesses and advertisers can better target audiences by region and content type.



Key points

Stress Detection Dashboard Report

Introduction

This project analyzes various lifestyle, behavioral, and demographic factors to understand their correlation with stress levels. Unlike previous projects where data cleaning and preprocessing were required, this dataset was already preprocessed, so the analysis directly focused on visualization and extracting insights.

The dashboard provides an overview of sleep quality, screen time, work hours, exercise, smoking habits, and social interactions, all of which contribute to stress detection.

Dataset Overview

- The dataset included variables such as:
  - Sleep quality & duration
  - Screen time
  - Social interaction score
  - Work hours
  - Physical activity (hrs)
  - Lifestyle habits (smoking, travel time, exercise)
  - Demographic factors (gender, occupation, marital status, age)
  - Stress detection levels: High, Medium, Low

Key Insights from Dashboard

1. General Lifestyle Metrics:
    - Average Sleep Quality: 3.85 (on a scale)
    - Average Screen Time: 4.1 hours/day
    - Average Social Interaction: 3.2 (scale-based)
    - Average Work Hours: 8.26 hours/day
    - Average Physical Activity: 3 hours/day
    - Average Sleep Duration: 6.3 hours/day
2. Smoking Habits:
    - Majority (≈ 62%) are non-smokers, while around 38% reported smoking.
3. Stress Level vs Screen Time:
    - Individuals with high stress tend to have higher screen time.
    - Screen time decreases in the low stress group but slightly rises again for medium stress.
4. Stress by Gender:
    - Males show slightly higher stress levels compared to females across all categories.
    - High stress: Male (23.16%) > Female (15.78%)
    - Low stress: Male (9.18%) < Female (11.77%)
5. Exercise vs Sleep Quality:
    - Strength training (31.69%) and Cardio (27.81%) participants reported higher sleep quality.
    - Minimal improvements were seen in Meditation and Walking.
6. Screen Time vs Marital Status:
    - Married individuals had the highest average screen time compared to single and divorced groups.
7. Travel Time Distribution:
    - Balanced between male and female participants:
        - Female: 50.32%
        - Male: 49.68%

Conclusion

The analysis shows that lifestyle habits such as screen time, exercise, and smoking play a crucial role in stress levels. Longer work hours and lower social interactions also contribute to stress. Males reported higher stress levels overall, while physical exercise (especially strength training and cardio) significantly improved sleep quality, a critical factor for stress reduction.

Since the dataset was already preprocessed, the focus remained entirely on insight generation and visualization. This dashboard can help organizations and individuals identify stress patterns and take corrective lifestyle actions.