**Weather and Entertainment Consumption Analysis**

**Abstract**

This project explores the relationship between temperature variations and personal entertainment consumption patterns. By analyzing data collected over a specified period, we investigate how changes in temperature influence the choice and duration of entertainment activities such as streaming, gaming, or outdoor leisure. Using exploratory data analysis and predictive modeling techniques, we uncover trends and insights into behavioral shifts associated with weather changes. The findings highlight key temperature thresholds that correlate with shifts in entertainment preferences, providing actionable insights for businesses and personal planning. Limitations and future research directions are also discussed.

**Introduction**

Understanding the factors that influence entertainment consumption is vital in today’s data-driven world, where businesses aim to align their offerings with consumer behavior. Among these factors, weather, and specifically temperature, plays a subtle but impactful role in shaping how individuals choose to spend their leisure time. This project investigates how variations in temperature affect entertainment consumption, leveraging personal data to identify trends and patterns. By combining exploratory data analysis with predictive modeling, the study provides insights into how preferences shift with changing weather conditions. These findings not only shed light on individual habits but also offer practical implications for industries such as streaming services and event planning.

**Data Description**

To explore the relationship between temperature and entertainment consumption, data was collected from three primary sources: Spotify, Netflix, and a weather API. These platforms provided information about streaming durations, type of content, and corresponding weather conditions.

**Spotify Data:**

Spotify usage data included details about the number of songs played and the total listening duration per day. This provided a comprehensive view of music consumption patterns.

**Netflix Data:**

Netflix data consisted of the titles and total viewing times for shows and movies streamed during the analysis period, offering insights into video content preferences.

**Weather Data:**

Weather data, obtained from a reliable API, included daily temperature, humidity, and other meteorological factors for the user’s location. Temperature was the primary focus variable for this study.

Despite initial efforts to include YouTube data, it was excluded from the analysis. The lack of information on content durations and the complexity of retrieving this data made its inclusion infeasible within the project timeline.

To ensure consistency, all data sources were synchronized by date, allowing for the analysis of consumption patterns alongside temperature variations. Data preprocessing steps included cleaning, normalization, and merging datasets from different sources.

A graph with orange and blue lines

Description automatically generated

A graph of a graph showing the temperature of a person

Description automatically generated with medium confidenceA graph with orange and blue lines

Description automatically generated

A graph with blue and orange dots

Description automatically generated

**Methodology**

To investigate the impact of temperature on entertainment consumption, a multi-step methodology was applied, encompassing data collection, preprocessing, model development, and evaluation.

**A. Data Collection**

Data was gathered from Spotify, Netflix, and a weather API:

Spotify provided streaming durations in minutes for each day.

Netflix offered daily viewing durations for movies and shows.

Weather API supplied meteorological data, with a focus on daily average temperatures.

**B. Data Preprocessing**

To ensure consistency and usability:

Dates in all datasets were converted to a standard format, and rows with invalid or missing dates were removed.

Datasets were merged using the common date column, allowing analysis of streaming data alongside temperature values.

Missing values in usage data were filled with zeros to maintain completeness for model training.

**C. Model Development**

Two models were developed to capture different aspects of the data:

**Classification Model:** Aimed to predict the platform preference (Netflix or Spotify) based on temperature using a Random Forest Classifier.

**Regression Models**: Separate Random Forest Regressors were trained to predict the time spent on Netflix and Spotify, respectively, as functions of temperature.

**D. Model Implementation**

Classification Model: A new categorical feature, "Platform," was created to label days where Netflix consumption exceeded Spotify (and vice versa).

Regression Models: Netflix and Spotify durations were converted into hours for improved interpretability.

**E. Model Training**

The dataset was split into training and testing subsets for both classification and regression tasks.

Random Forest models were trained using the training data with a focus on generalization.

**F. Model Evaluation**

The classification model was evaluated using metrics like accuracy and a classification report.

The regression models were assessed using the Mean Squared Error (MSE) for predicted vs. actual durations.

**G. Data Aggregation and Visualization**

Temperature ranges were defined, and average consumption for Netflix and Spotify was computed within each range.

A comparative analysis was visualized, highlighting how usage patterns varied across temperature bands.

**Challenges Encountered**

During preprocessing, complications such as date mismatches required significant effort to resolve. These issues limited the scope of exploratory data analysis, particularly in generating comprehensive visual insights.

**Findings**

The analysis revealed key insights into how temperature affects entertainment consumption patterns:

**Platform Preferences:**

On colder days (average temperature below 10°C), Netflix usage was notably higher than Spotify usage, suggesting a preference for indoor activities.

As temperatures increased (above 20°C), Spotify consumption showed an upward trend, indicating a shift towards portable entertainment suitable for outdoor environments.

**Duration Trends:**

Netflix consumption peaked during moderate to cold weather conditions, particularly around the 10–20°C range, with average durations exceeding 2 hours per day.

Spotify usage showed a gradual increase with rising temperatures, reaching its highest average daily usage in temperatures above 30°C.

**Impact of Temperature:**

Regression models showed that temperature significantly influenced the time spent on both platforms, with Netflix usage negatively correlated and Spotify usage positively correlated with temperature.

**Limitations Observed:**

Data preprocessing challenges, such as date mismatches, restricted the depth of exploratory data analysis and may have introduced some noise in the merged dataset.

Exclusion of YouTube data limited the overall scope of the analysis.

**Conclusion**

This study demonstrated that temperature plays a significant role in shaping entertainment consumption patterns. Colder weather drives users towards indoor, screen-based entertainment like Netflix, while warmer temperatures encourage music streaming through platforms like Spotify, likely due to its compatibility with outdoor activities.

Despite some limitations in data and analysis, the findings provide valuable insights for businesses in the entertainment industry, enabling them to better align their offerings with seasonal consumer behavior. Future work could incorporate additional entertainment platforms, refine data collection methods, and explore other weather factors like humidity and precipitation to further understand the relationship between environmental conditions and entertainment preferences.