PW SKILLS

Assignment

EDA

Amazon Sales Data

Description:

This dataset contains information on 1K+ Amazon products, including their ratings, reviews, and other details.

Features:

product_id: Unique identifier for each product

product_name: Name of the product

category: Category of the product

discounted_price: Discounted price of the product

actual_price: Actual price of the product

discount_percentage: Percentage of discount for the product

rating: Rating of the product (1-5)

rating_count: Number of people who voted for the Amazon rating

about_product: Description about the product

user_id: ID of the user who wrote the review

user_name: Name of the user who wrote the review

review_id: ID of the user review

review_title: Short review

review_content: Long review

img_link: Image link of the product

product_link: Official website link of the product

Source: Amazon Sales

Questions:

1. What is the average rating for each product category?
2. What are the top rating_count products by category?
3. What is the distribution of discounted prices vs. actual prices?

4. How does the average discount percentage vary across categories?

5. What are the most popular product names?

6. What are the most popular product keywords?

7. What are the most popular product reviews?

8. What is the correlation between discounted_price and rating?

9. What are the Top 5 categories based on the highest ratings?

10. Identify any potential areas for improvement or optimization based on the data analysis.

Dataset Link: **Spotify Data:Popular Hip-Hop Artists and Tracks**🎶

Description of the Dataset:

The dataset titled "Spotify Data: Popular Hip-hop Artists and Tracks" provides a curated collection of approximately 500 entries showcasing the vibrant realm of hip-hop music. These entries meticulously compile the most celebrated hip-hop tracks and artists, reflecting their significant influence on the genre's landscape. Each entry not only highlights the popularity and musical composition of the tracks but also underscores the creative prowess of the artists and their profound impact on global listeners.

Application in Data Science:

This dataset serves as a valuable resource for various data science explorations. Analysts can delve into trend analysis to discern the popularity dynamics of hit hip-hop tracks over recent years. Additionally, the dataset enables network analysis to uncover collaborative patterns among top artists, shedding light on the genre's evolving collaborative landscape. Furthermore, it facilitates the development of predictive models aimed at forecasting track popularity based on diverse features, offering insights for artists, producers, and marketers.

Column Descriptors:

Artist: The name of the artist, providing direct attribution to the creative mind behind the track.

Track Name: The title of the track, encapsulating its identity and essence.

Popularity: A numeric score reflecting the track's reception and appeal among Spotify listeners.

Duration (ms): The track's length in milliseconds, detailing the temporal extent of the musical experience.

Track ID: A unique identifier within Spotify's ecosystem, enabling direct access to the track for further exploration.

Questions:

1.Load the dataframe and ensure data quality by checking for missing values and duplicate rows.
Handle missing values and remove duplicate rows if necessary.

2.What is the distribution of popularity among the tracks in the dataset? Visualize it using a histogram.

3.Is there any relationship between the popularity and the duration of tracks? Explore this using a scatter plot.

4.Which artist has the highest number of tracks in the dataset? Display the count of tracks for each artist using a countplot.

5.What are the top 5 least popular tracks in the dataset? Provide the artist name and track name for each.

6.Among the top 5 most popular artists, which artist has the highest popularity on average?
Calculate and display the average popularity for each artist.

7.For the top 5 most popular artists, what are their most popular tracks? List the track name for each artist.

8.Visualize relationships between multiple numerical variables simultaneously using a pair plot.

9.Does the duration of tracks vary significantly across different artists?
Explore this visually using a box plot or violin plot.

10.How does the distribution of track popularity vary for different artists?
Visualize this using a swarm plot or a violin plot.