**Birla Institute of Technology & Science, Pilani**
**Work Integrated Learning Programmes Division**
**Second Semester 2022-2023**

**Mid-Semester Test**
**(EC-2 Makeup)**

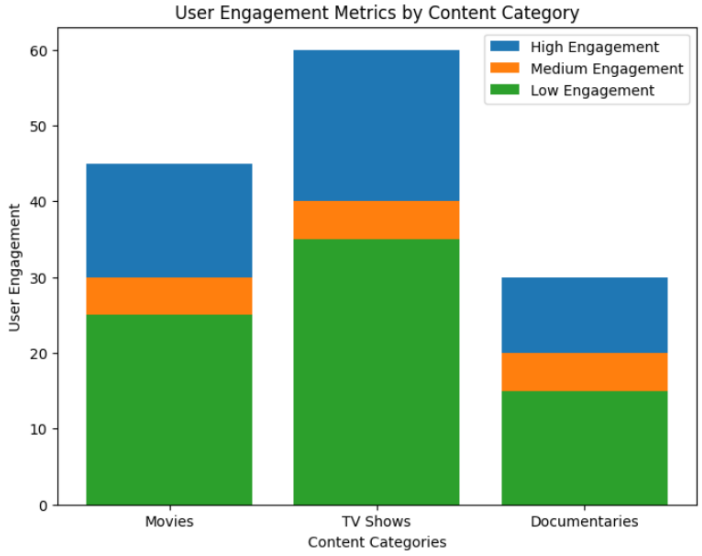
| | |
|---|---|
| Course No. | : DSECLZG555 |
| Course Title | : Data Visualization and Interpretation |
| Nature of Exam | : Open Book |
| Weightage | : 30% |
| Duration | : 2 Hours |
| Date of Exam | : 22 July 2023 (FN) |

No. of Pages = 03
No. of Questions = 05

Note to Students:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

**Please Note**: Choice of chart type, Decluttering, Effective use of Gestalt principles, Strategic use of pre-attentive attributes, Logical ordering, and effective use of text -contribute towards the credit. Most effective chart fetches maximum credit. [Pen and paper drawing would suffice. No additional credit for usage of any Viz Tools for creating visuals]

**Q.1.** The visual provided below illustrates the user engagement metrics for different content categories using a stacked bar chart. Each content category, namely Movies, TV Shows, and Documentaries, is divided into three engagement levels: High Engagement, Medium Engagement, and Low Engagement. Answer the following questions regarding the visualization context: **[4 Marks]**




**A.** **WHO** could be the intended audience for this data visualization? [1 mark]
**B.** **WHAT** is being represented in the stacked bar chart? [1 mark]
**C.** **HOW** does the visualization effectively convey the user engagement metrics for different content categories? [1 mark]
**D.** Propose a **Big idea.** [1 mark]

**Answer key**: Possible solution:

**A.** The intended audience for this data visualization could be **content creators**, media companies, or streaming platforms. It could also be **marketing or analytics teams** interested in understanding user engagement trends related to different content categories. [Full marks (1M) if the right audience is identified]

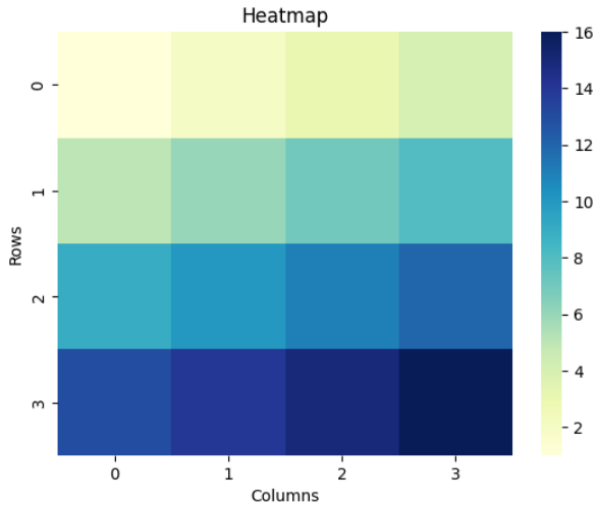
**B.** The stacked bar chart **represents the user engagement metrics for different content categories**. Each bar represents a content category (Movies, TV Shows, and Documentaries), and each bar is divided into three segments representing the levels of engagement: High Engagement, Medium Engagement, and Low Engagement. [Full marks (1M) if chart representation explained rightly]

**C.** The visualization effectively **conveys the user engagement metrics for different content categories by using a stacked bar chart**. The use of different colors for each engagement level allows viewers to quickly distinguish between high, medium, and low engagement for each content category. Additionally, the chart's stacked nature allows viewers to see the total engagement for each content category as well as the distribution of engagement levels within each category. This makes it easy to compare engagement levels between different content categories and identify which category has the highest or lowest engagement. [Full marks (1M) if the 'how' part is explained]

**D.**

**E.** **Big Idea**: "Maximizing User Engagement Through Content Category Optimization" [Full marks (1M) if the big idea is stated as one single statement concisely]

**Q.2.** How does the color intensity in the heatmap represent the values of the data? Interpret the heatmap below and discuss any patterns or trends that can be observed based on the color variations. **[4 Marks]**



**Answer key**: Possible solution:

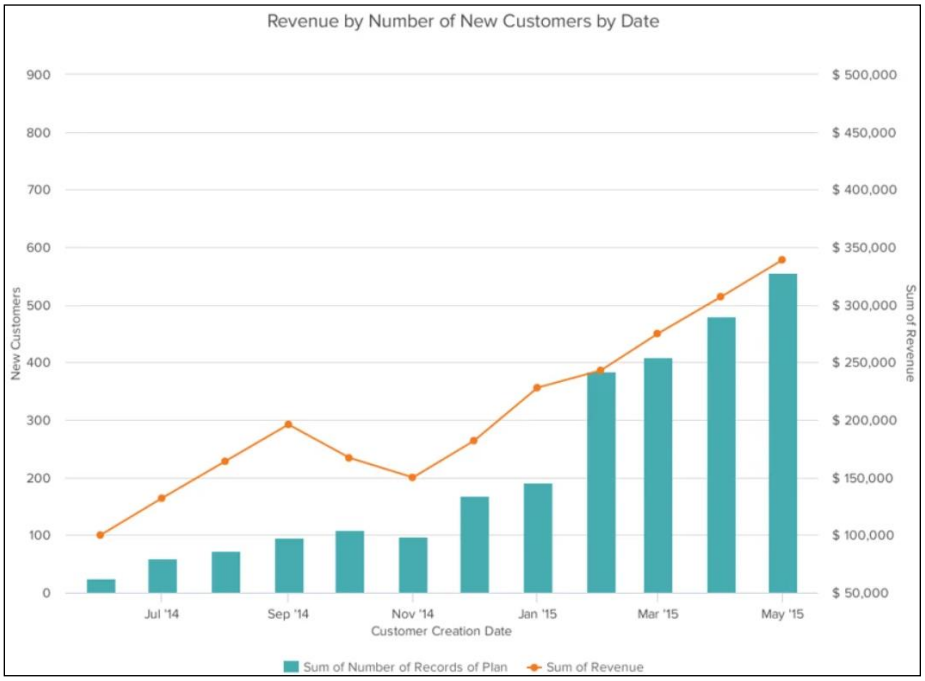
In a heatmap, color intensity represents the values of the data through a color gradient. Typically, a darker or more intense color indicates a higher value, while a lighter or less intense color indicates a lower value. The specific color scale used in the heatmap may vary, but the general principle is to use a spectrum of colors to represent the range of data values.

[Full marks (4M) if explained considering the points mentioned below]

- Color variations and patterns
- High engagement
- Low engagement
- Comparisons
- Clusters and correlations
- Outliers

**Q.3.** The given dual axis chart illustrates the revenue and the number of new customers over time for a company. The line graph represents the revenue, while the bar chart represents the number of new customers. Each data point on the chart corresponds to a specific date.

Discuss the limitations and drawbacks of using a dual axis chart to visualize the relationship between revenue and the number of new customers. Justify your reasons for avoiding this type of chart (3 marks). Furthermore, propose and create two alternative visuals that could serve as more effective replacements for the dual axis chart, considering the clarity and accuracy of data representation (3 marks). **[6 Marks]**



**Answer key:** Possible solution:
[Full marks (6M) if explained considering points below, (3M) for limitations and drawbacks of dual axis chart and (3M) for alternative visual sketch]

- Limitations
  - Misleading comparisons
  - Visual clutter
  - Loss of precision
- Alternative visuals
  - Simple line chart (or)
  - Grouped bar chart (or) any other effective visual

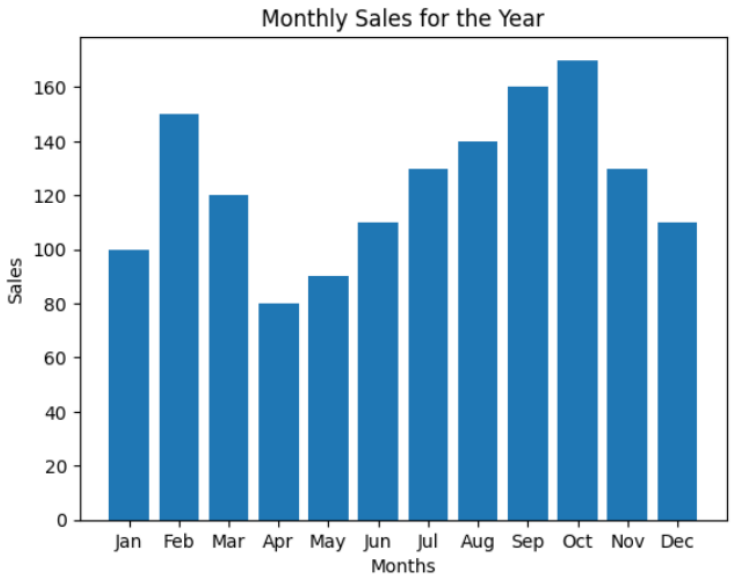
**Q.4.** Based on the given two-way table, which represents movie preferences by gender, **propose and draw the four most effective visuals** to analyze and represent the data. Justify your choices and explain how each visual can provide meaningful insights into movie preferences based on gender.

| | **Action** | **Comedy** | **Drama** | **Horror** |
|---|---|---|---|---|
| **Male** | 20 | 15 | 10 | 5 |
| **Female** | 10 | 20 | 15 | 8 |

The visuals should effectively convey the analysis of movie preferences, allowing for a clear understanding of the gender-based preferences for different movie genres. **[8 Marks]**

**Answer key:** Possible solution:
[Full marks (8M) if 4 different (2M each – 1M for graph and 1M for justification)]
For instance, the 4 alternative visuals could be, stacked bar chart, grouped bar chart, heatmap, line chart.

**Q.5.** The bar chart provided below represents the monthly sales data for a year, ranging from January to December. Discuss the potential improvements that could enhance the **visualization context using gestalt principles and pre-attentive attributes**, making it more effective in conveying the monthly sales trends over the course of the year **[8 Marks]**



**Answer key:** Possible solution:

[Full marks (8M) if 6 improvements are suggested] The possible solution would be something similar to,

1. Sorting bars
2. Color coding
3. Adding totals
4. Minimal annotations
5. Focal point – highlighting the focal point
6. Trend line
7. Scaling/labeling
8. Final visual (2M)

**********