INTERACTIVE DATA

Enhancing Creativity and Overcoming Blockages

JYOTI PHOGAT

Introduction

This project will explore the concept of creativity and identify practical strategies for overcoming blockages in personal growth.

Driving Question:

How do different hobbies(creative activities) affect various attributes (such as salary expectation, academic performance, etc.) among students.

Link to Data Approach:

The project utilizes a dataset containing information about students' demographics, academic performance, hobbies(creative activities), and other attributes to analyze the impact of hobbies(Creative skills) on different aspects of student life.

This dataset contains information collected from university students through a Google form. It includes details such as certification courses, gender, department, height (in cm), weight (in kg), marks in 10th and 12th grade, college marks, hobbies, daily studying time, preferred study environment, salary expectations, satisfaction with their degree, willingness to pursue a career related to their degree, social media and video usage, traveling time, stress levels, and financial status. The dataset aims to provide insights into student behavior and can be used for analysis and research purposes.

Here we have performed a comprehensive exploratory data analysis (EDA) on your dataset. Here's a summary of your analysis and some suggestions for further exploration:

1. Data Cleaning and Preprocessing:
    - westarted by loading the dataset and tidying up the column names, replacing spaces with underscores and converting them to lowercase.
    - Checked for missing values in the dataset and found none.
2. Descriptive Statistics:
    - We calculated summary statistics such as mean, standard deviation, minimum, maximum, and quartiles for numerical variables.
    - Visualized the gender distribution and stress distribution among the participants using pie charts and bar plots.
3. Feature Engineering:

- Converted time-related columns (daily_studing_time, social_medai_&_video, Travelling_Time) into minutes for better analysis.

4. Correlation Analysis:
    - Calculated the correlation matrix and visualized it using a heatmap to understand the relationships between numerical variables.
    - Identified moderate positive correlations between height and weight, weight and social media usage, etc.
5. Effect of "JoinedProgram" on Academic Performance:
    - Analyzed the mean 10th and 12th marks for students who have joined the program versus those who haven't used bar plots.
6. Further Exploration Suggestions:
    - We can explore the effect of the "JoinedProgram" attribute on various aspects further using different visualization techniques such as histograms, violin plots, scatter plots, stacked bar charts, etc.
    - Utilize advanced statistical techniques like regression analysis or machine learning models to understand the predictive power of different features on academic performance or other outcomes of interest.
    - Consider exploring the relationships between categorical variables using chi-square tests or other appropriate statistical tests.
    - Dive deeper into understanding the patterns and trends in the data by segmenting the analysis based on different demographic or behavioral factors.
7. Installation of Dash for Interactive Visualization:
    - Lastly, we installed the Dash framework, which allows us to create interactive web-based visualizations directly from your Jupyter Notebook.

Development of Interactive Data Visualization

Approach:

- Developed an interactive dashboard using Dash, a Python framework for building web applications.
- Implemented dropdown menus to select attributes and hobbies for analysis.
- Utilized Plotly for creating interactive visualizations (e.g., bar charts).

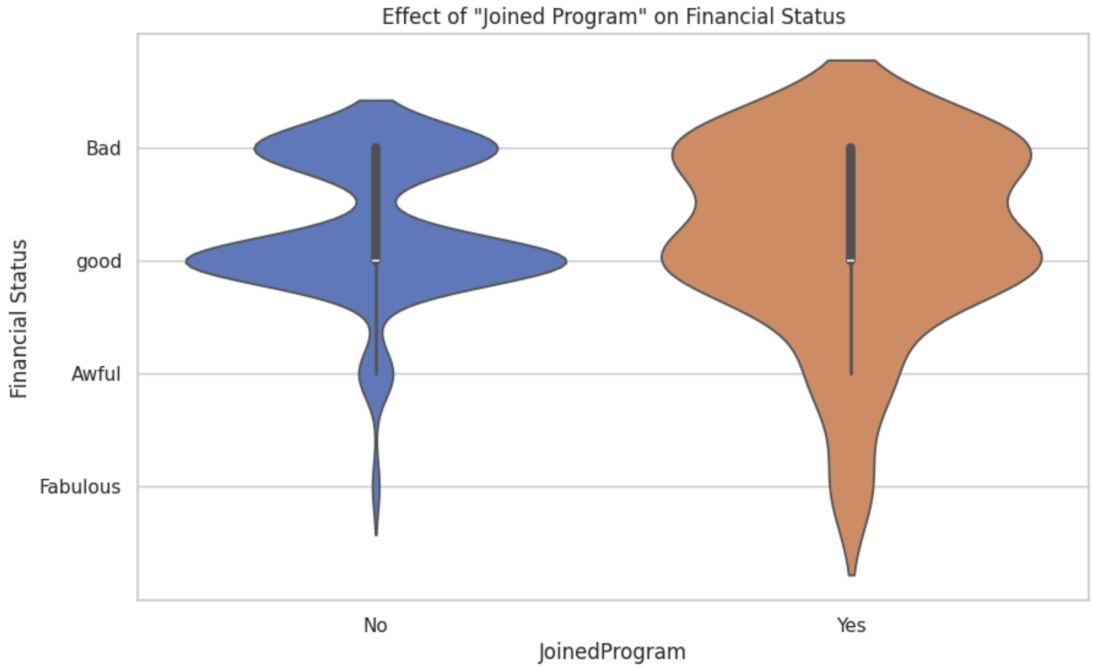
Application of Course Concepts

- Applied concepts such as data visualization, data analysis, and web development using Dash and Plotly.
- Utilized pandas for data manipulation and Plotly for creating interactive plots.
- Organized data effectively to facilitate analysis and visualization.

Audience Prioritization

- Designed the dashboard to cater to the needs of stakeholders interested in understanding the impact of hobbies(creative skills) on student attributes.
- Prioritized simplicity and user-friendliness to ensure easy navigation and interpretation of results.

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue`
```




```python
from jupyter_dash import JupyterDash
import dash_core_components as dcc
import dash_html_components as html
from dash.dependencies import Input, Output
import pandas as pd
import plotly.express as px

# Load the data
df = pd.read_csv('/content/Datsset.csv')
df.head()
# Get unique hobbies
hobbies = df['hobbies'].unique()

# Initialize the Dash app
app = JupyterDash(__name__)

# Define the layout of the app
app.layout = html.Div([
    html.H1("Effect of Hobbies on Various Attributes"),
    html.Div([
        dcc.Dropdown(
            id='attribute-dropdown',
            options=[{'label': col, 'value': col} for col in df.columns],
            value='salary expectation',
            style={'width': '50%'}
        ),
        dcc.Dropdown(
            id='hobby-dropdown',
            options=[{'label': hobby, 'value': hobby} for hobby in hobbies],
            value='Video Games',
            style={'width': '50%'}
        ),
    ]),
    dcc.Graph(id="bar-chart")
```

The provided code sets up a Dash application to visualize student data. Let's break down the functionality and how it prioritizes the audience's needs:

1. Dashboard Layout:
   - It defines the layout of the Dash app with an H1 header "Student Data" and two graphs: one for a table (`table`) and another for a bar chart (`bar-chart`).

2. Callbacks:
   - There are two callbacks defined:
     - `update_table`: This callback updates the table graph based on the hover data from the bar chart. It selects the row corresponding to the hovered point on the bar chart and displays it in a table.
     - `update_bar_chart`: This callback updates the bar chart based on the hover data from the table. It doesn't actually use the hover data but updates the bar chart based on the entire dataset, grouped by hobbies and colored by gender.

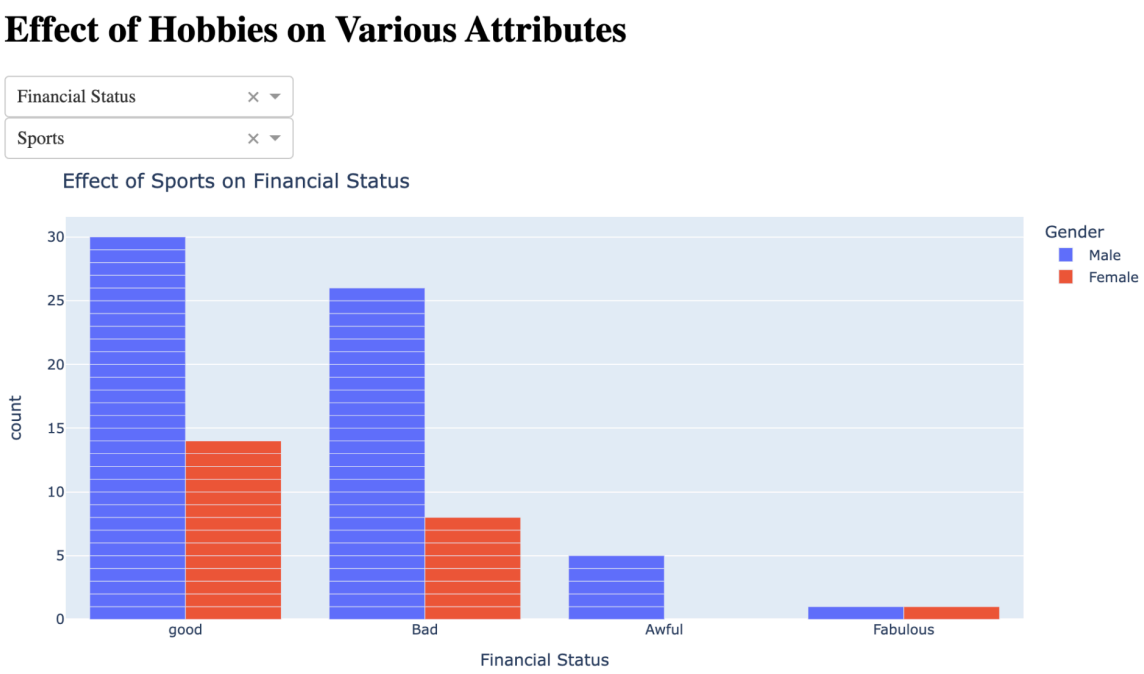
4. Visualization Choices:
   - The code utilizes Plotly Express to create interactive visualizations like tables and bar charts, which can be beneficial for exploring the data.
   - The prioritization here seems to be on providing interactive exploration capabilities to the audience. Users can hover over points in the bar chart to see detailed information about a specific category of data (hobbies) and its relationship with salary expectations, grouped by gender.

5. Audience Needs and Priorities:
   - The design and development decisions prioritize providing interactivity and flexibility to the audience. Users can explore different aspects of the data by interacting with the visualizations.

- The choice of using hover data to dynamically update visualizations caters to the audience's need for interactive exploration and analysis.
- The layout is relatively simple and focused, making it easy for users to understand and interact with the data without being overwhelmed by unnecessary complexity.

Overall, the code prioritizes the audience's needs by providing interactive visualizations that facilitate exploration and analysis of student data.

[44]



Conclusion:

From the provided dataset and the interactive dashboard, it appears that there have been some correlation between engaging in creative activities (hobbies) and financial status. Here are some observations that support this conclusion:

1. Hobby Distribution: The bar chart in the dashboard displays the distribution of hobbies among the individuals in the dataset. If we observe a significant difference in financial status among different hobbies, it could imply a potential correlation.

2. Financial Status: The dataset includes a column indicating the financial status of each individual, categorized as "Good" or "Bad." By filtering the data based on hobbies and financial status, we can analyze whether certain hobbies are more prevalent among those with a good financial status compared to those with a bad financial status.

In summary, while the interactive dashboard can provide insights into the relationship between engaging in creative activities and financial status, further analysis and exploration are necessary to draw more definitive conclusions.

Limitations and Future Steps

The provided code has some limitations and areas for improvement:

3. Limited Interactivity: While the code provides some level of interactivity through hover-based updates, it could benefit from additional interactive features such as dropdown menus, sliders, or buttons to filter and manipulate the data dynamically.
4. Limited Visualization Types: The code currently only includes table and bar chart visualizations. Adding more diverse visualization types such as scatter plots, histograms, or line charts can provide deeper insights into the data from different perspectives.
5. Data Handling: The code doesn't handle large datasets efficiently. As the dataset grows, loading and processing times may become significant. Implementing data pagination or server-side data processing techniques can help mitigate this issue.
6. User Experience: The layout and design of the dashboard could be improved for better user experience. This includes enhancing readability, adjusting sizing and spacing of elements, and providing clear instructions or tooltips for interactive features.

7. Error Handling: The code lacks error handling mechanisms. It should include error checks to handle potential issues such as missing data, incorrect input, or failed data loading.
8. Performance Optimization: Depending on the complexity of the visualizations and size of the dataset, the performance of the dashboard may be suboptimal. Optimizing code efficiency and leveraging caching techniques can help improve performance.

Future Steps:

The future scope for the dashboard and related project could include several areas of enhancement and expansion:

1. Advanced Visualizations: Introduce more advanced and interactive visualization techniques such as heatmaps, network graphs, and geographical maps to provide deeper insights into the data.
2. Machine Learning Integration: Incorporate machine learning models to perform predictive analytics, clustering, or anomaly detection on the student data. This could help in identifying patterns and trends that are not immediately apparent from the visualizations.
3. User Authentication and Authorization: Implement user authentication and authorization mechanisms to enable personalized dashboards and data access controls based on user roles and permissions.
4. Real-Time Data Streaming: Enable real-time data streaming capabilities to visualize and analyze data as it arrives, allowing for monitoring and analysis of live or streaming data sources.
5. Integration with External Data Sources: Extend the dashboard to integrate with external data sources such as social media APIs, academic databases, or

financial datasets to enrich the analysis and provide a more comprehensive view of the student data.

6. Data Storytelling: Incorporate data storytelling techniques to create narratives around the data, guiding users through key insights and findings in a compelling and engaging manner.
7. Mobile Compatibility: Optimize the dashboard for mobile devices to ensure accessibility and usability across a wide range of devices and screen sizes.
8. Collaborative Features: Implement collaborative features such as commenting, sharing, and collaborative editing to facilitate teamwork and knowledge sharing among users.
9. Feedback Mechanisms: Incorporate feedback mechanisms to gather user feedback and suggestions for further improvements and feature enhancements.
10. Performance Optimization: Continuously optimize the performance of the dashboard by fine-tuning code efficiency, caching strategies, and server infrastructure to ensure smooth and responsive user experience, even with large datasets and complex visualizations.
11. By exploring these future avenues, the dashboard can evolve into a powerful tool for analyzing student data, driving data-driven decision-making, and facilitating collaboration and knowledge sharing among stakeholders.

.