UDACITY

Return to Classroom

Investigate a Dataset

REVIEW

HISTORY

Meets Specifications

Dear Student,

Congratulations ⭐
You have done great job on updating the project, it reflects your dedication for learning. Excellent work on raising good questions and discussing the insights in the conclusion with limitations 👏🏽

Some suggestions and links are provided for additional learning. If you have any questions, feel free to post them on knowledge. Stay Udacious

Some additional links for learning:

- Data exploration
- How to create effective visualizations

Code Functionality

- **All code is functional and produces no errors when run.**
- **The code given is sufficient to reproduce the results described.**

All code is functional and produces no errors when run. The code given is sufficient to reproduce the results described.

Some resources for learning:

- Commonly Pandas functions
- Python
- Guide For Markdown cells

- **The project uses NumPy arrays and Pandas Series and DataFrames where appropriate rather than Python lists and dictionaries.**
- **Where possible, vectorized operations and built-in functions are used instead of loops.**

The project correctly uses Pandas series and dataframes throughout the project. Good use of inbuilt functions has been made.

- **The code makes use of at least 1 function to avoid repetitive code.**
- **The code contains good comments and meaningful variable names, making it easy to read.**

The notebook has well commented code, the variable names are apt and related to the project. Good work on adding function to avoid repetitive code.

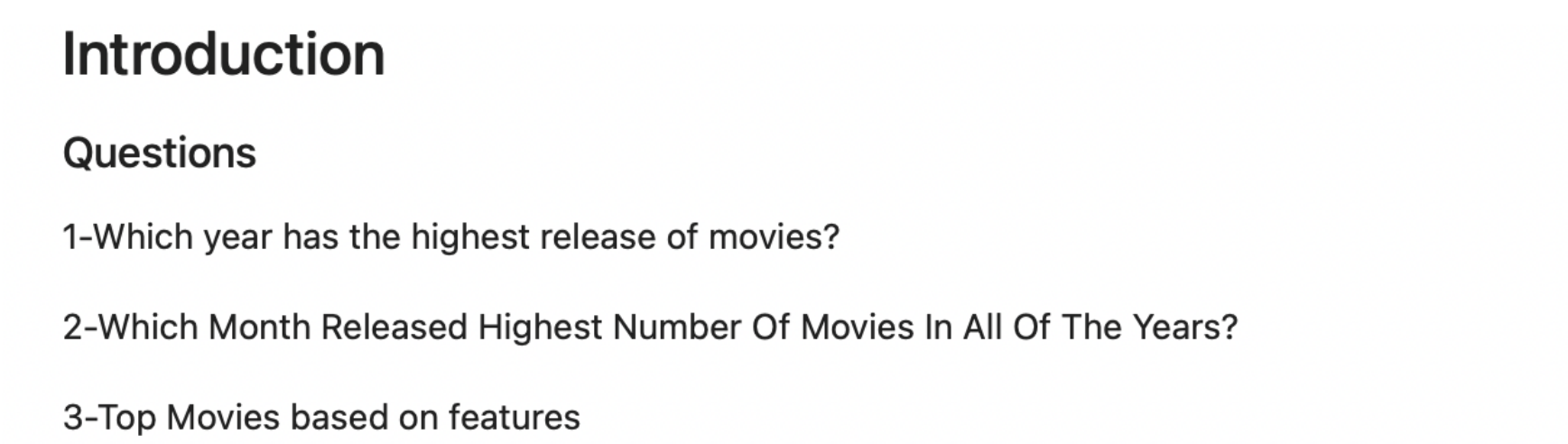
```
def drawFig(title):
    # Remove axes splines
    for s in ['top', 'bottom', 'left', 'right']:
        ax.spines[s].set_visible(False)


    # Remove x, y Ticks
    ax.xaxis.set_ticks_position('none')
    ax.yaxis.set_ticks_position('none')

    # Add padding between axes and labels
    ax.xaxis.set_tick_params(pad = 5)
    ax.yaxis.set_tick_params(pad = 10)

    # Add x, y gridlines
    ax.grid(visible = True, color ='grey', linestyle ='-.', linewidth = 0.5, alpha = 0.2)

    # Show top values
    ax.invert_yaxis()

    # Add annotation to bars
    for i in ax.patches:
        plt.text(i.get_width()+0.2, i.get_y()+0.5,
                 str(round((i.get_width()), 2)),
                 fontsize = 10, fontweight ='bold',
                 color ='grey')
```

Quality of Analysis

The project clearly states one or more questions, then addresses those questions in the rest of the analysis

The project clearly states one or more questions, then addresses those questions in the rest of the analysis.

The project has clearly raised more than one good specific questions and addressed them later on.



Data Wrangling Phase

The project documents any changes that were made to clean the data, such as merging multiple files, handling missing values, etc.

Suggestion - Although previous reviewer missed this, but the 0 values in budget, revenue and runtime columns should be handled as missing values in data cleaning. I would encourage you to read this article which suggests some ways to handle the null values in the data

Exploration Phase

- The project investigates the stated question(s) from multiple angles.
- The project explores at least three variables in relation to the primary question. This can be an exploratory relationship between three

variables of interest, or looking at how two independent variables relate to a single dependent variable of interest.
- The project performs both single-variable (1d) and multiple-variable (2d) explorations.

The project correctly investigates the questions mentioned from multiple angles. These visualizations are relevant to the questions asked in the project.

- The project's visualizations are varied and show multiple comparisons and trends.
- At least two kinds of plots should be created as part of the explorations.
- Relevant statistics are computed throughout the analysis when an inference is made about the data.

There is no limit to the visualizations that can be done for a project, but I can see that you have more than two plots as part of EDA. Good job on the visualizations.

Conclusions Phase

- The Conclusions have reflected on the steps taken during the data exploration.
- The Conclusions have summarized the main findings in relation to the question(s) provided at the beginning of the analysis accurately.
- The project has pointed out where additional research can be done or where additional information could be useful.
- The conclusion should have at least 1 limitation explained clearly.
- The analysis does not state or imply that one change causes another based solely on a correlation.

Good job on providing clear limitation

Conclusions

1- Year 2014 year has the highest release of movies (700) followed by year 2013 (659) and year 2015 (629).

2- The higher number of release in september and october month.

3- The top 5 movies from the given dataset based on their adjusted revenue are the followings; Avatar, Star Wars, Titanic, The Exorcist and Jaws.

4- The top 5 movies from the given dataset based on their adjusted budget are the followings; The Warrioi's Way, Pirates of the Caribbean. On Stranger Tides, Pirates of the Caribbean. At World's End, Superman Returns and Titanic.

5- The top 5 movies from the given dataset based on their Popularity are the followings; Jurassic World, Mad Max: Fury Road, Interstellar, Gaurdians of the Galaxy and Insurgent.

6- The top 5 movies from the given dataset based on their Voting Average are the followings; The Story of Film: An Odyssey, Black Mirror: White Christmas, Pink Floyd: Pulse, The Art of Flight and A Personal journey with Martin scoursese Through American Movies.

The limitations associated with the conclusions are:

There is a big limitation here, as can be seen from data processed above, around 52 % of budget data is zero !! which affects profit calculation greatly. Combined with zero revenue, around 65 % of profit is zero or wrongly calculated !!

Communication

- The code should have ideally the following sections: Introduction; Questions; Data Wrangling; Exploratory Data Analysis; Conclusions, Limitation.
- Reasoning is provided for each analysis decision, plot, and statistical summary.
- Interpretation of plots and application of statistical tests should be correct and without error.
- Comments are used within the code cells.
- Documented the flow of analysis in the mark-down cells.

Correct reasoning has been given for the visualizations and statistical summary made in the notebook

Visualizations made in the project depict the data in an appropriate manner (i.e., has appropriate labels, scale, legends, and plot type) that allows plots to be readily interpreted.

The plots are provided a title, a clear mention of labels on X & Y axis. Good use of legends has been made 💯

DOWNLOAD PROJECT

RETURN TO PATH