UDACITY

< Return to Classroom

Investigate a Dataset

REVIEW

HISTORY

Meets Specifications

Congratulations!! 🎉🎉

> Excellent work! **Incorporating with the previous feedback .**

Now , Your submission has passed the rubric of this project.

All the effort you put in to complete the project is appreciated and it was my pleasure reviewing your work.

> ## Extra Useful Material for functions in programming 💡
>
> - Why use functions in programming?
> - See this material for the importance of using functions in python.

.

Good luck in your educational progress 😄

.

Code Functionality

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

Well done!

==========

- The **submitted code works well** as it doesn't produce errors when run.
- Also, it's **sufficient to reproduce the results** described.

.

- 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.

.

Excellent Work

==========

- The analysis makes use of the **NumPy and Pandas libraries,** vector operators are employed instead of loops and lists.
- It is awesome that you make use of the function `.info()`, `.describe()`, `groupby()`, `value_counts()`, `.duplicated()` to examine the structure of the entire data, identify missing values, and the summary statistics for the numerical features.

Extra Material

==========

Some important Pandas built-in functions:

• Group-by: http://pandas.pydata.org/pandas-docs/stable/groupby.html

• Value-Counts: https://chrisalbon.com/python/data_wrangling/pandas_dataframe_count_values/

.

- 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.

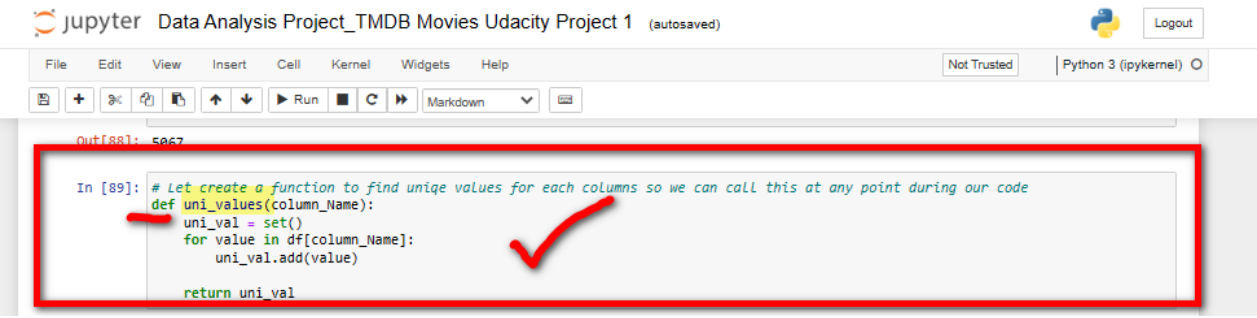
.

Good work

==========

- Using **descriptive names and comments** in your code makes it easier for other programmers to follow up on the work.

.



.

Learning Notes

==========

- Why use functions in programming?
- See this material for the importance of using functions in python.

.

Quality of Analysis

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

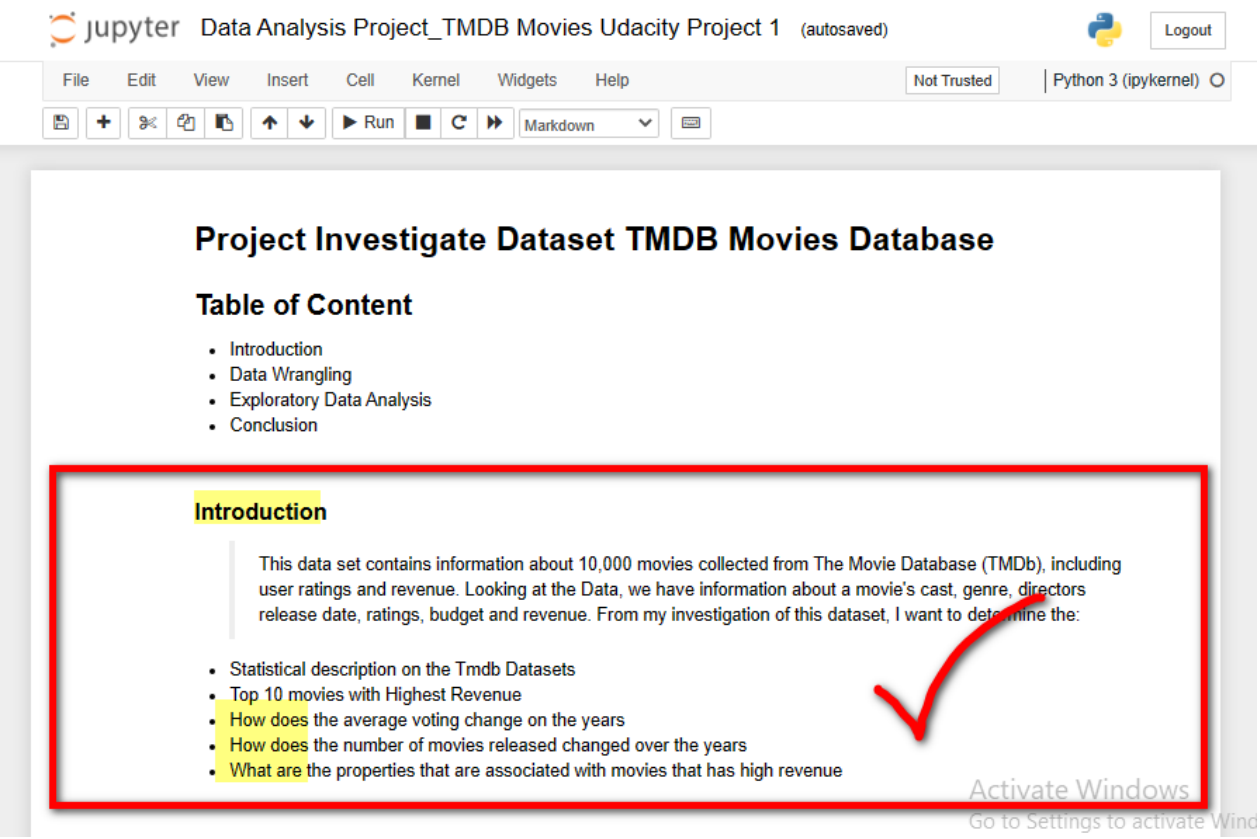
.

Good job

==========

- In the **introduction section,** You made a brief description of the project dataset.
- You **list the questions** right after the introduction, so the readers could catch your points in the first place, **Well done**
- In the **EDA section,** you answer the questions with plots.

.



Data Wrangling Phase

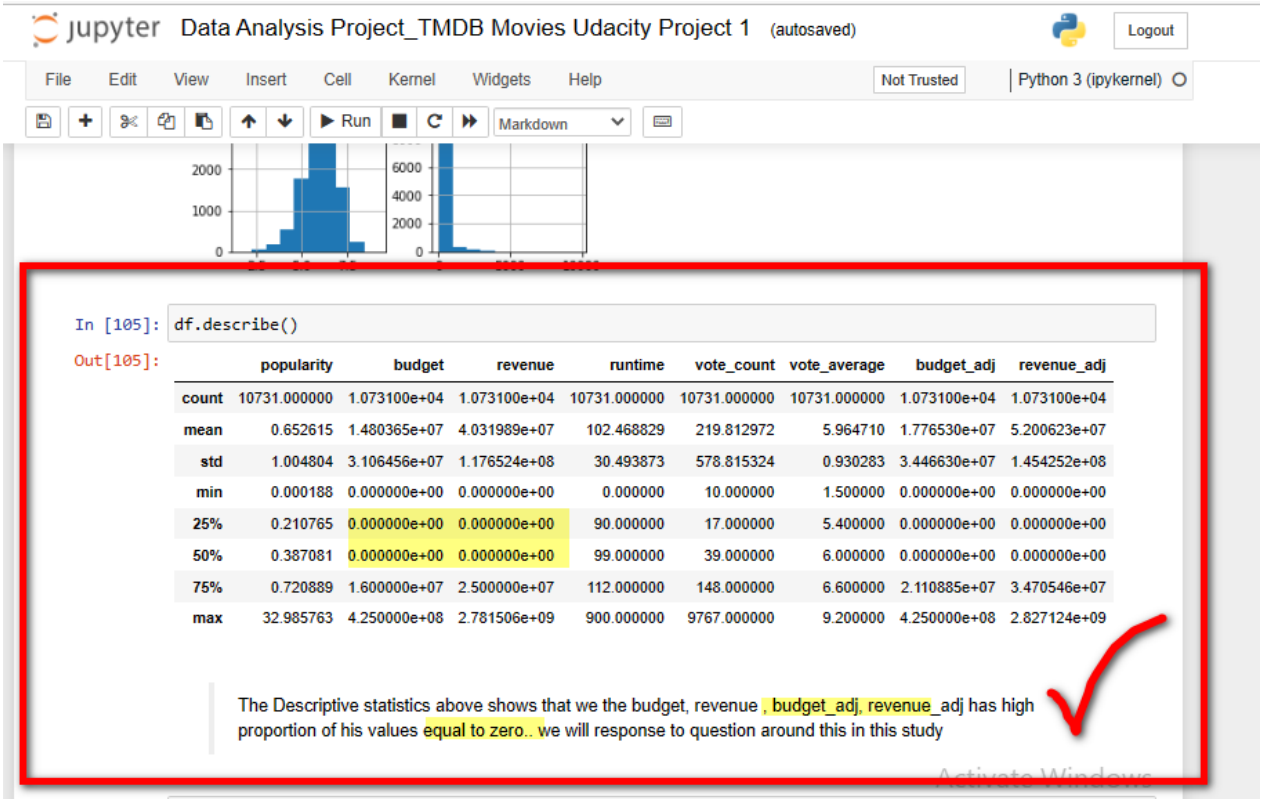
The project documents any changes that were made to clean the data, such as merging multiple files, handling missing values, etc.

Great job

==========

- Your notebook is **clean and well-structured.**
- You added **detail** by using **appropriate** `comments and many markdown texts` to explain **what** your code does and **how** it works.
- **Different sections** are clearly shown the data wrangling process.
- You did a perfect data wrangling, you have looked into the **numeric statistic, data type, missing value, and duplicates**
- You identify the **outlier of numeric column's value** `zero values` and investigate it, e.g **revenue.**

.



	popularity	budget	revenue	runtime	vote_count	vote_average	budget_adj	revenue_adj
count	10731.000000	1.073100e+04	1.073100e+04	10731.000000	10731.000000	10731.000000	1.073100e+04	1.073100e+04
mean	0.652615	1.480365e+07	4.031989e+07	102.468829	219.812972	5.964710	1.776530e+07	5.200623e+07
std	1.004804	3.106456e+07	1.176524e+08	30.493873	578.815324	0.930283	3.446630e+07	1.454252e+08
min	0.000188	0.000000e+00	0.000000e+00	0.000000	10.000000	1.500000	0.000000e+00	0.000000e+00
25%	0.210765	0.000000e+00	0.000000e+00	90.000000	17.000000	5.400000	0.000000e+00	0.000000e+00
50%	0.387081	0.000000e+00	0.000000e+00	99.000000	39.000000	6.000000	0.000000e+00	0.000000e+00
75%	0.720889	1.600000e+07	2.500000e+07	112.000000	148.000000	6.600000	2.110885e+07	3.470546e+07
max	32.985763	4.250000e+08	2.781506e+09	900.000000	9767.000000	9.200000	4.250000e+08	2.827124e+09

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.

.

Good job

==========

- The analysis makes use of both **single and multiple variables** explorations to investigate different features and the relations between these features in the dataset.

.

- **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.**

.

Good job

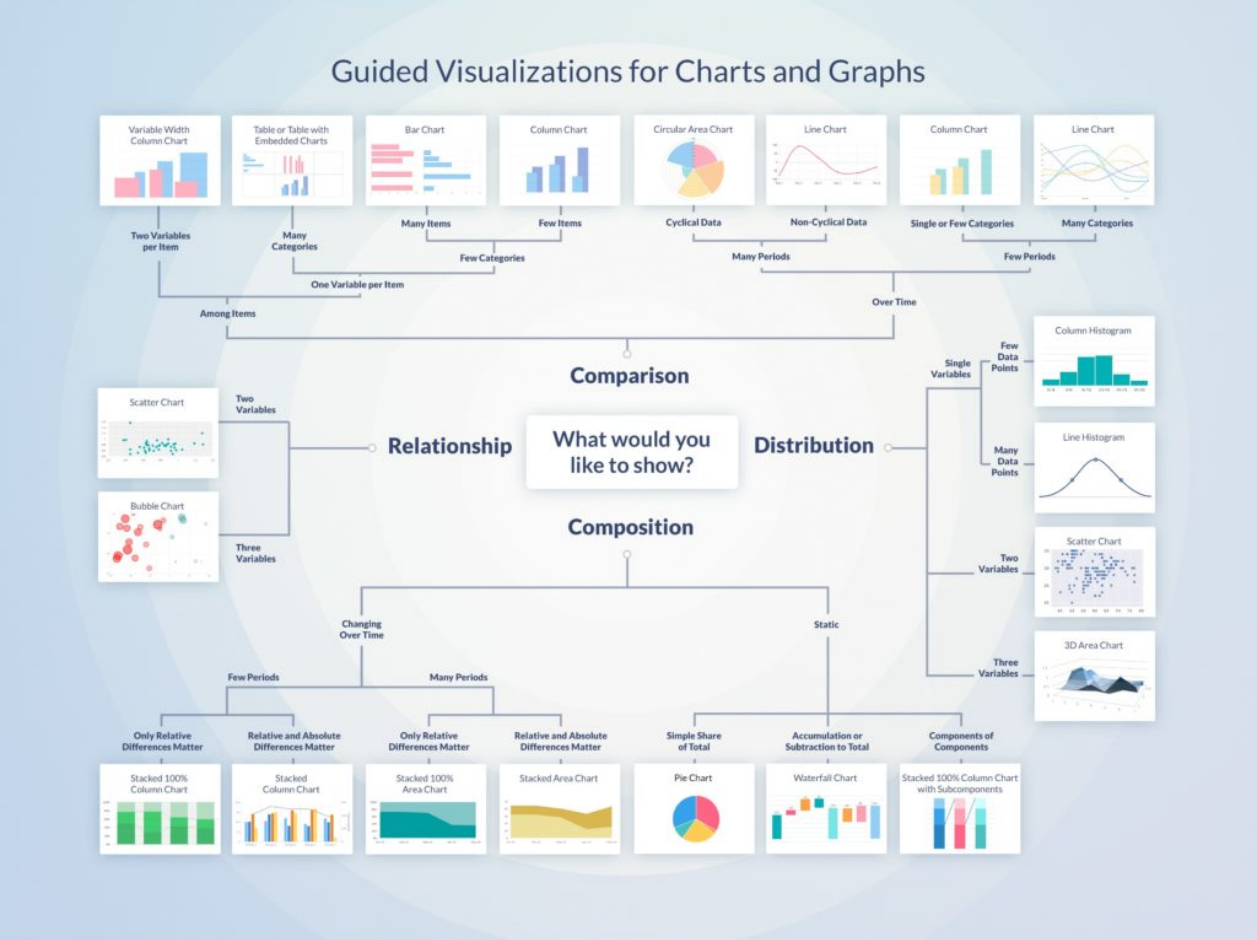
==========

- **Multiple plots** have been properly used, great job!

> ## Extra Material

Below is a guideline for graphing, just for your future reference.

.



.

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.

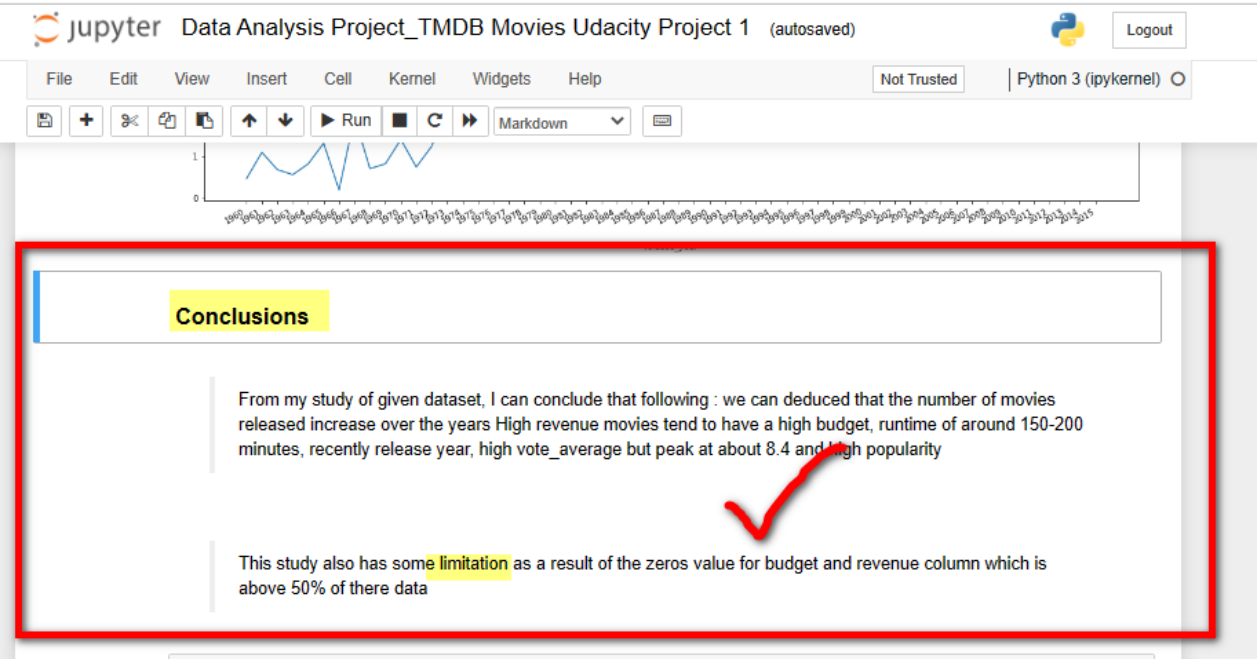
.

Good work

==========

- Presenting the **results** of the analysis while showing its **limitations** clearly.

.



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.

Great job

==========

- **Describing every analysis decision**, and plot stating the **results** obtained.

.

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.

Great job

==========

- The plots have **titles and labels** and are easy to interpret.

.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

START