UDACITY

< Return to Classroom

Wrangle and Analyze Data

REVIEW

HISTORY

Meets Specifications

Congratulations!! 🎉 🎉

Excellent work! 👍 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 Materials

Data visualization is one of the most valuable skills for data scientists/analysts, **So** we recommended you to know more about :

- **Different kinds of charts with python seaborn library**, you can take a look at it in this link.
- **Also** the following link gives a good guideline to use what chart.

.
Good luck with your educational progress 😊

.

Code Functionality and Readability

All project code is contained in a Jupyter Notebook named wrangle_act.ipynb and runs without errors.

> ### Good job
>
> ==========
>
> - **Files** submitted correctly, and your notebook **doesn't** show any execution **errors.**

.

The Jupyter Notebook has an intuitive, easy-to-follow logical structure. The code uses comments effectively and is interspersed with Jupyter Notebook Markdown cells. The steps of the data wrangling process (i.e. gather, assess, and clean) are clearly identified with comments or Markdown cells, as well.

> ### Well done
>
> ==========
>
> - The project notebook is **clean and well-structured.**
> - **Different sections** are clearly shown for each one of the steps of the data wrangling process.
> - **Also you** added detail **by using appropriate comments** that were needed to explain **what** your code does and **how** it works.

.

Gathering Data

Data is successfully gathered:

- From at least the three (3) different sources on the Step 1: Gathering Data page.
- In at least the three (3) different file formats on the Step 1: Gathering Data page.

Each piece of data is imported into a separate pandas DataFrame at first.

.

Good work

==========

- Data had been successfully gathered from **three different sources.**
- Each piece of data was imported into a **separate object at first.**

.

Assessing Data

Two types of assessment are used:

- Visual assessment: each piece of gathered data is displayed in the Jupyter Notebook for visual assessment purposes. Once displayed, data can additionally be assessed in an external application (e.g. Excel, text editor).
- Programmatic assessment: pandas' functions and/or methods are used to assess the data.

> ## Well done
>
> ==========
>
> - The **visual assessment** has been done.
> - Programmatic assessment
>
> The project used pandas/python functions to programmatically assess the dataset using `info()` , `value_counts()` , `query()` , `.describe()` , `sum()` , `duplicated()` and other useful functions to explore the data.
>
> .

> ## 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/

.

At least eight (8) data quality issues and two (2) tidiness issues are detected, and include the issues to clean to satisfy the Project Motivation. Each issue is documented in one to a few sentences each.

> ## Good job

==========

- You have identified **data quality and tidiness issues** and explained the action that should be taken to them.
- You have **given a clear and succinct explanation** for each one of them.

> ## Note 💡

==========

Just in case keep in mind, that there is no definitive list, there are definitely other different points to consider, but you have identified the most important issues

.

Cleaning Data

The define, code, and test steps of the cleaning process are clearly documented.

.

Well done

==========

- You have correctly **performed the cleaning data step** of the data wrangling process.
- The **Define - Code - Test** framework has been correctly applied.

.

Copies of the original pieces of data are made prior to cleaning.

All issues identified in the assess phase are successfully cleaned (if possible) using Python and pandas, and include the cleaning tasks required to satisfy the Project Motivation.

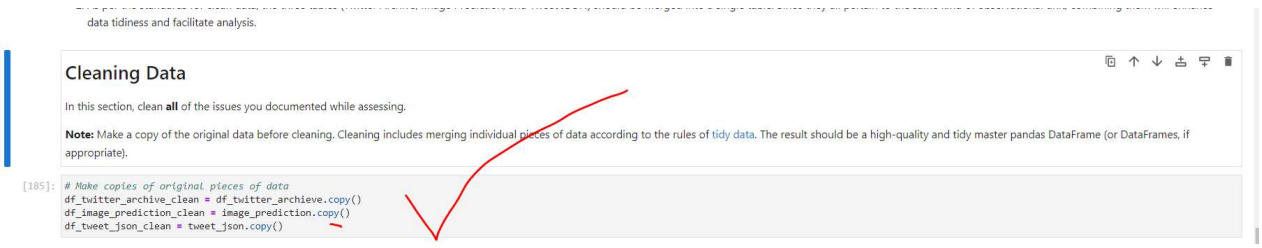
A tidy master dataset (or datasets, if appropriate) with all pieces of gathered data is created.

> ## Great job!

==========

- You **copied the original** data before cleaning.
- Also, you created the **final clean master** dataset.

.



Storing and Acting on Wrangled Data

Students will save their gathered, assessed, and cleaned master dataset(s) to a CSV file or a SQLite database.

.

Meets Specification

==========

- The **gathered, assessed, and cleaned** master dataset(s) are saved to a **CSV file.**

.

The master dataset is analyzed using pandas or SQL in the Jupyter Notebook and at least three (3) separate insights are produced.

At least one (1) labeled visualization is produced in the Jupyter Notebook using Python's plotting libraries or in Tableau.

Students must make it clear in their wrangling work that they assessed and cleaned (if necessary) the data upon which the analyses and visualizations are based.

Good job

==========

- The master dataset is analyzed using **the pandas** and insights and visualizations are given.
- The visualization is produced in the Jupyter Notebook using **Python's plotting libraries.**

.

.

Suggestion :

==========

- It is very easy to plot different kinds of charts with the **python Seaborn library.**
- You can look at this link for more details about the Python Seaborn library , as Data visualization is one of the most valuable skills for data scientists/analysts.

.

Report

The student's wrangling efforts are briefly described. This document (wrangle_report.pdf or wrangle_report.html) is concise and approximately 300-600 words in length.

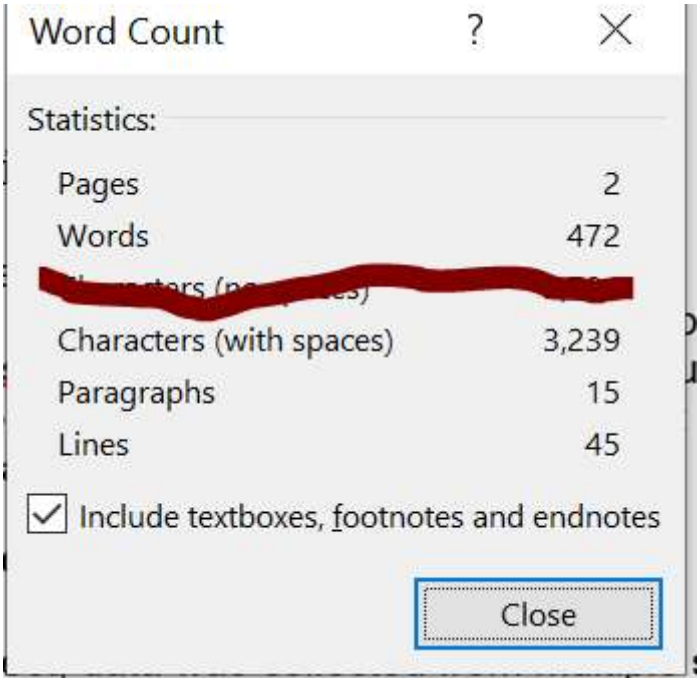
.

Good job

==========

- Creating the **report with wrangling** efforts.
- It's **clear and concise and reflects the wrangling process** taken for the data set.
- This document **(wrangle_report)** is between 300-600 words in length.

.



The three (3) or more insights the student found are communicated. At least one (1) visualization is included.

This document (act_report.pdf or act_report.html) is at least 250 words in length.

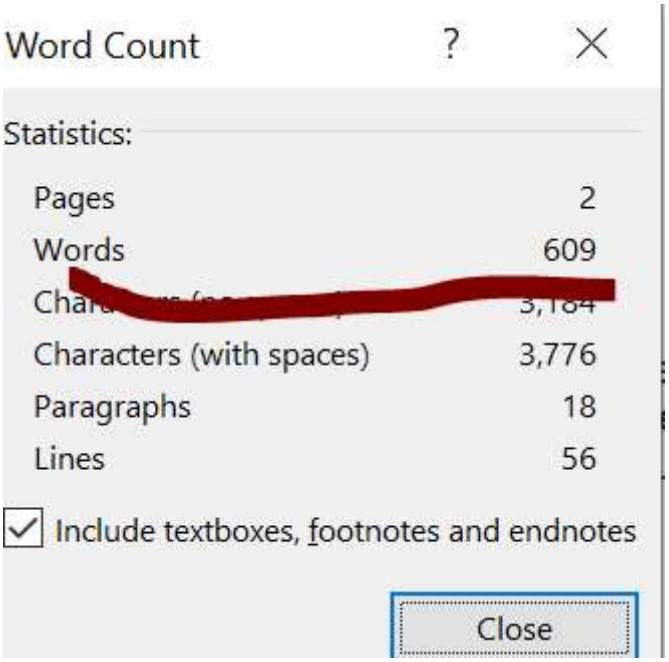
.

Good work!

==========

- Included **many** visualizations in your `act_report` **Report .**
- Creating **many** insights in your `act_report` **Report .**
- This document included more than **250 words** in length.

.



Project Files

The following files (with identical filenames) are included:

- wrangle_act.ipynb
- wrangle_report.pdf or wrangle_report.html
- act_report.pdf or act_report.html

All dataset files are included, including the stored master dataset(s), with filenames and extensions as specified on the Project Submission page.

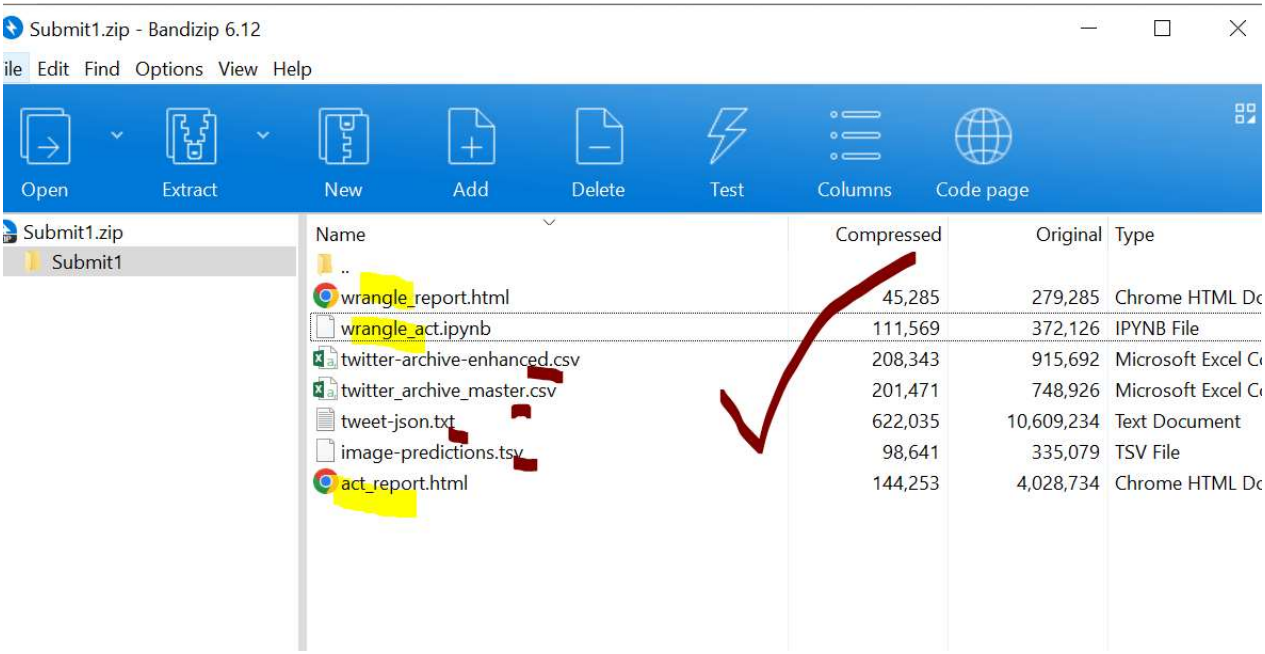
.

Good job

==========

- Added the required files in a proper format.

.



DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

START