UDACITY

< Return to Classroom

DISCUSS ON STUDENT HUB

Write a Data Science Blog Post

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Awesome 👏👏

Congratulations on passing this project. 🏆🏆

This project was not easy but you have done it gracefully. This is all because of your hard work and continuous evaluation. But still don't relax, keep exploring and learning from the references provided. Try to read more related articles on CRISP-DM. Try to solve and practice this kind of data analysis on Kaggle. This will bring a habit and will develop a mental model.

Extra Materials

The links below, provide more insights to these processes:

- CRISP-DM Overview- Data Science Project Management
- Manage your next Data Science Project with CRISP-DM METHOD
- Six steps in CRISP-DM the standard data mining process

You can also share your project on LinkedIn and ask the audience for necessary feedback or open the project for anyone to collaborate. This way you will find many interesting connections and engagement with others.

I wish you good luck. Looking forward to your success.
For any queries, you can ask on Knowledge Portal as well.

Stay ! Stay Safe

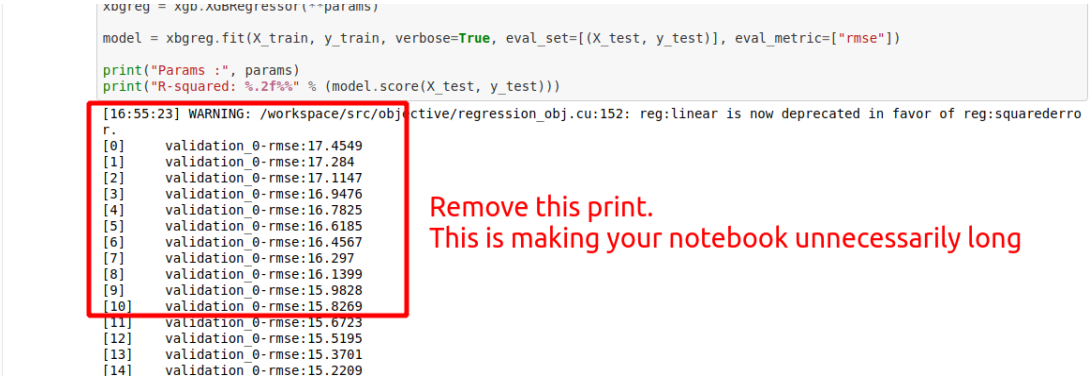
DON'T FORGET TO RATE MY WORK AS PROJECT REVIEWER! YOUR FEEDBACK IS VERY HELPFUL AND APPRECIATED.

Code Functionality and Readability

All the project code is contained in a Jupyter notebook, which demonstrates successful execution and output of the code.

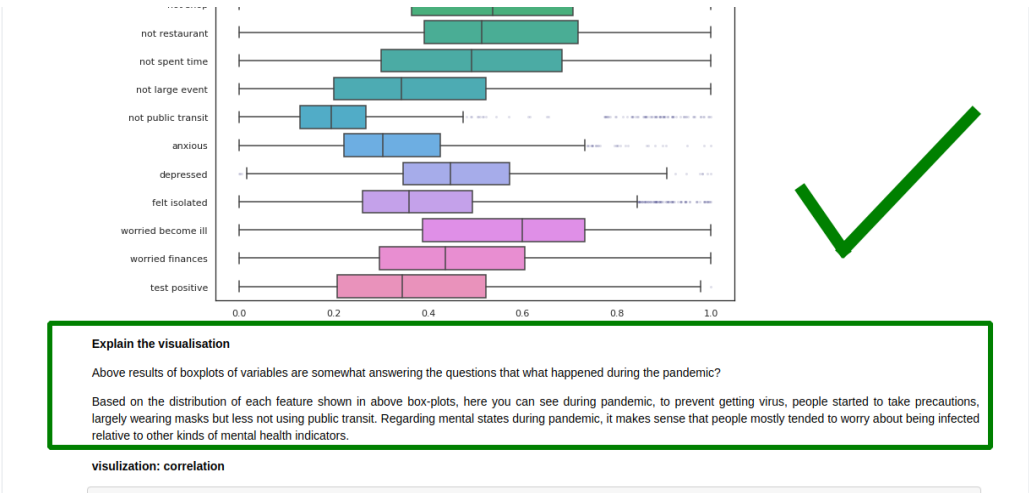
The provided code in this submission shows successful execution without errors. 💯

Suggestion:

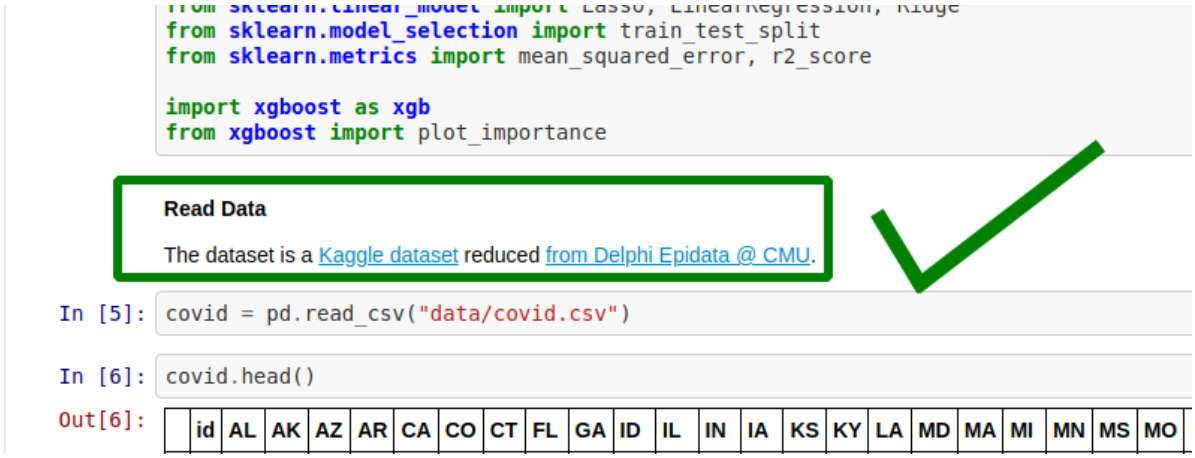


Code has easy-to-follow logical structure. The code uses comments effectively and/or Notebook Markdown cells correctly. The steps of the data science process (gather, assess, clean, analyze, model, visualize) are clearly identified with comments or Markdown cells, as well. The naming for variables and functions should be according to PEP8 style guide.

✅ Code has easy-to-follow logical structure. The code uses comments effectively and/or Notebook Markdown cells correctly.



✅ The steps of the data science process (gather, assess, clean, analyze, model, visualize) are clearly identified with comments or Markdown cells, as well



```
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

import xgboost as xgb
from xgboost import plot_importance
```

Read Data

The dataset is a Kaggle dataset reduced from Delphi Epidata @ CMU.

```
In [5]: covid = pd.read_csv("data/covid.csv")

In [6]: covid.head()
```

Out[6]:

	id	AL	AK	AZ	AR	CA	CO	CT	FL	GA	ID	IL	IN	IA	KS	KY	LA	MD	MA	MI	MN	MS	MO

✅ The naming for variables and functions should be according to PEP8 style guide.

Useful References:

Jupyter Notebook Best Practices

Jupyter notebook shortcuts

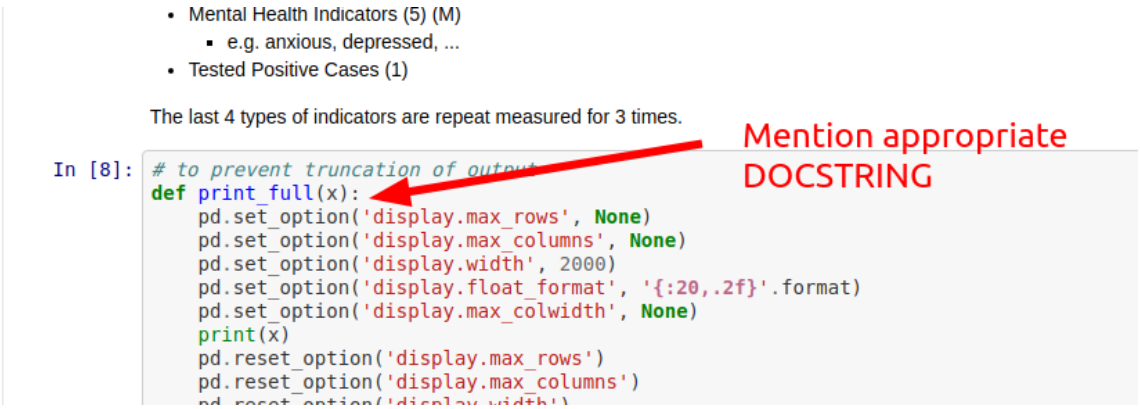
PEP 8 -- Style Guide for Python Code

Code is well documented and uses functions and classes as necessary. All functions include document strings. DRY principles are implemented.

Passed by previous reviewer

✅ Code is well documented and uses functions and classes as necessary.

✅ All functions include document strings



- Mental Health Indicators (5) (M)
  - e.g. anxious, depressed, ...
- Tested Positive Cases (1)

The last 4 types of indicators are repeat measured for 3 times.

Mention appropriate DOCSTRING

```
In [8]: # to prevent truncation of output
def print_full(x):
    pd.set_option('display.max_rows', None)
    pd.set_option('display.max_columns', None)
    pd.set_option('display.width', 2000)
    pd.set_option('display.float_format', '{:20,.2f}'.format)
    pd.set_option('display.max_colwidth', None)
    print(x)
    pd.reset_option('display.max_rows')
    pd.reset_option('display.max_columns')
```

```
    pd.reset_option('display.width')
    pd.reset_option('display.float_format')
    pd.reset_option('display.max_colwidth')

print_full(covid_dtypes)
```

✅ DRY principles are implemented.

Useful References:

Python Docstrings

Docstring vs Comments

Do not Repeat Yourself

Data

Project follows the CRISP-DM process outlined for questions through communication. This can be done in the README or the notebook. If a question does not require machine learning, descriptive or inferential statistics should be used to create a compelling answer to a particular question.

Your analysis does an exceptional job of following the CRISP-DM process. It is clear what part of the process you are on in each part of the notebook, which is difficult to do in a notebook. Nice job!

Categorical variables are handled appropriately for machine learning models (if models are created). Missing values are also handled appropriately for both descriptive and ML techniques. Document why a particular approach was used, and why it was appropriate for a particular situation.

As data is already preprocessed, so didn't get the chance to apply or follow any approach to handle.

But if you do in the future, then make sure that you have followed right approach to fill them or to drop them or even to keep them. And also try to explain why you have followed a certain approach in the markdown cells.

Further Readings:

- Simple Methods to deal with Categorical Variables in Predictive Modeling
- Missing Values in Data - Statistics Solutions
- How to Handle Missing Data -- Towards Data Science

Analysis, Modeling, Visualization

In the Jupyter Notebook, there are between 3-5 questions asked, related to the business or real-world context of the data. Each question is answered with appropriate visualization, table, or statistic.

Nice job with coming with your business questions. You have some really interesting questions with some cool findings. You have nicely put forward your hypotheses with support of great visual. Well Done!

Github Repository

Student must have a Github repository of their project. The repository must have a README.md file that communicates the libraries used, the motivation for the project, the files in the repository with a small description of each, a summary of the results of the analysis, and necessary acknowledgements. Students should not use another student's code to complete the project, but they may use other references on the web including StackOverflow and Kaggle to complete the project.

Passed by previous reviewer

✅ Student must have a Github repository of their project.

✅ communicates the libraries used

✅ motivation for the project

✅ files in the repository

✅ summary of the results of the analysis

❌ necessary acknowledgments

Provide an acknowledgement section and acknowledge the dataset and reference used in the project.

Useful References:

Manage your data science project structure in early stage

How to write a good readme for your github project?

Blog Post

Student must have a blog post on a platform of their own choice (can be on their website, a Medium post or Github blog post). The post should not dive into technical details or difficulties of the analysis - this should be saved for Github. The post should be understandable for non-technical people from many fields.

Awesome work on the blog post.

I really liked how you structure your analysis. Your post clearly communicates findings to non-technical stakeholders. You did a good job of providing a clear structure from the intro, the questions of interest, and a conclusion.

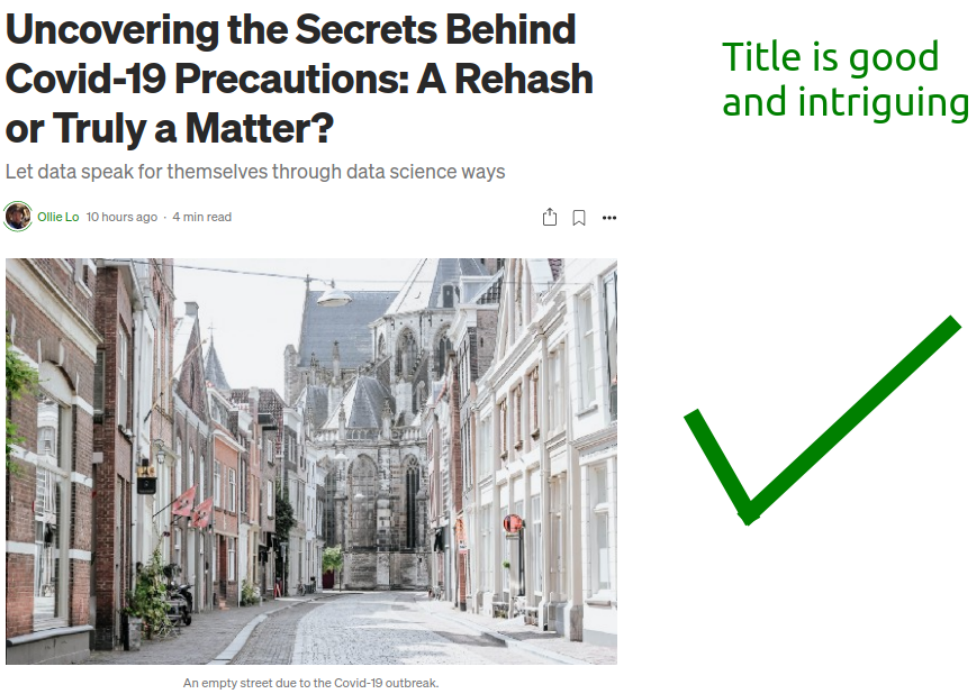
Useful References:

A blog post from the instructor

Art of Storytelling

5 reasons why programmers should blog

Student must have a title and image to draw readers to their post.



Useful References:

Medium Titles, Subtitles, and Kickers

How you write headlines

There are no long, ongoing blocks of text without line breaks or images for separation anywhere in the post.

You did a great job of making sure your post was broken into short paragraphs with clear ideas in each.

Each question is clearly stated and each answer includes a clear visual, table, or statistic.

Each question was followed with a visual of what provided a potential solution to the question. The discussion around these plots is really well done too!

What happened during the pandemic?

Based on the distribution of each feature shown in above box-plots, here you can see during pandemic, to prevent getting virus, people started to take precautions, ***largely wearing masks*** but ***less not using public transit***. Regarding mental states during pandemic, it makes sense that people mostly tended to ***worry about being infected*** relative to other kinds of mental health indicators.

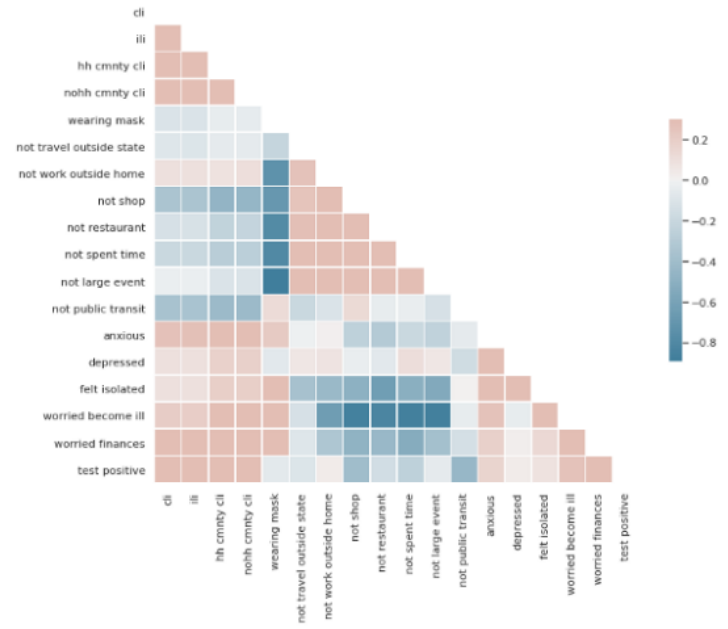


Figure 2: Correlations of Covid-related Features

Nice structure maintained throughout the blog post

DOWNLOAD PROJECT

RETURN TO PATH