

Data Analysis with Python: Zero to Pandas - Course Project Guidelines

(remove this cell before submission)

Important links:

- Make submissions here: <https://jovian.ml/learn/data-analysis-with-python-zero-to-pandas/assignment/course-project>
- Ask questions here: <https://jovian.ml/forum/t/course-project-on-exploratory-data-analysis-discuss-and-share-your-work/11684>
- Find interesting datasets here: <https://jovian.ml/forum/t/recommended-datasets-for-course-project/11711>

This is the starter notebook for the course project for [Data Analysis with Python: Zero to Pandas](#). You will pick a real-world dataset of your choice and apply the concepts learned in this course to perform exploratory data analysis. Use this starter notebook as an outline for your project. Focus on documentation and presentation - this Jupyter notebook will also serve as a project report, so make sure to include detailed explanations wherever possible using Markdown cells.

Evaluation Criteria

Your submission will be evaluated using the following criteria:

- Dataset must contain at least 3 columns and 150 rows of data
- You must ask and answer at least 4 questions about the dataset
- Your submission must include at least 4 visualizations (graphs)
- Your submission must include explanations using markdown cells, apart from the code.
- Your work must not be plagiarized i.e. copy-pasted for somewhere else.

Follow this step-by-step guide to work on your project.

Step 1: Select a real-world dataset

- Find an interesting dataset on this page: <https://www.kaggle.com/datasets?fileType=csv>
- The data should be in CSV format, and should contain at least 3 columns and 150 rows
- Download the dataset using the [opendatasets Python library](#).

Here's some sample code for downloading the [US Elections Dataset](#):

```
import opendatasets as od
dataset_url = 'https://www.kaggle.com/tunguz/us-elections-dataset'
od.download('https://www.kaggle.com/tunguz/us-elections-dataset')
```

You can find a list of recommended datasets here: <https://jovian.ml/forum/t/recommended-datasets-for-course-project/11711>

Step 2: Perform data preparation & cleaning

- Load the dataset into a data frame using Pandas
- Explore the number of rows & columns, ranges of values etc.
- Handle missing, incorrect and invalid data
- Perform any additional steps (parsing dates, creating additional columns, merging multiple dataset etc.)

Step 3: Perform exploratory analysis & visualization

- Compute the mean, sum, range and other interesting statistics for numeric columns
- Explore distributions of numeric columns using histograms etc.
- Explore relationship between columns using scatter plots, bar charts etc.
- Make a note of interesting insights from the exploratory analysis

Step 4: Ask & answer questions about the data

- Ask at least 4 interesting questions about your dataset
- Answer the questions either by computing the results using Numpy/Pandas or by plotting graphs using Matplotlib/Seaborn
- Create new columns, merge multiple dataset and perform grouping/aggregation wherever necessary
- Wherever you're using a library function from Pandas/Numpy/Matplotlib etc. explain briefly what it does

Step 5: Summarize your inferences & write a conclusion

- Write a summary of what you've learned from the analysis
- Include interesting insights and graphs from previous sections
- Share ideas for future work on the same topic using other relevant datasets
- Share links to resources you found useful during your analysis

Step 6: Make a submission & share your work

- Upload your notebook to your [Jovian.ml](https://jovian.ml) profile using `jovian.commit`.
- **Make a submission here:** <https://jovian.ml/learn/data-analysis-with-python-zero-to-pandas/assignment/course-project>
- Share your work on the forum: <https://jovian.ml/forum/t/course-project-on-exploratory-data-analysis-discuss-and-share-your-work/11684>
- Browse through projects shared by other participants and give feedback

(Optional) Step 7: Write a blog post

- A blog post is a great way to present and showcase your work.
- Sign up on [Medium.com](https://medium.com) to write a blog post for your project.
- Copy over the explanations from your Jupyter notebook into your blog post, and [embed code cells & outputs](#)
- Check out the [Jovian.ml](https://medium.com/jovianml) Medium publication for inspiration: <https://medium.com/jovianml>

Example Projects

Refer to these projects for inspiration:

- [Analyzing StackOverflow Developer Survey Results](#)
- [Analyzing Covid-19 data using Pandas](#)
- [Analyzing your browser history using Pandas & Seaborn](#) by Kartik Godawat
- [WhatsApp Chat Data Analysis](#) by Prajwal Prashanth
- [Understanding the Gender Divide in Data Science Roles](#) by Aakanksha N S
- [2019 State of Javascript Survey Results](#)
- [2020 Stack Overflow Developer Survey Results](#)

NOTE: Remove this cell containing the instructions before making your submission. You can do using the "Edit > Delete Cells" menu option.

Project Title - change this

TODO - Write some introduction about your project here: describe the dataset, where you got it from, what you're trying to do with it, and which tools & techniques you're using. You can also mention about the course [Data Analysis with Python: Zero to Pandas](#), and what you've learned from it.

How to run the code

This is an executable [Jupyter notebook](#) hosted on [Jovian.ml](#), a platform for sharing data science projects. You can run and experiment with the code in a couple of ways: *using free online resources* (recommended) or *on your own computer*.

Option 1: Running using free online resources (1-click, recommended)

The easiest way to start executing this notebook is to click the "Run" button at the top of this page, and select "Run on Binder". This will run the notebook on [mybinder.org](#), a free online service for running Jupyter notebooks. You can also select "Run on Colab" or "Run on Kaggle".

Option 2: Running on your computer locally

1. Install Conda by [following these instructions](#). Add Conda binaries to your system PATH, so you can use the `conda` command on your terminal.
2. Create a Conda environment and install the required libraries by running these commands on the terminal:

```
conda create -n zerotopandas -y python=3.8
conda activate zerotopandas
pip install jovian jupyter numpy pandas matplotlib seaborn opendatasets --upgrade
```

3. Press the "Clone" button above to copy the command for downloading the notebook, and run it on the terminal. This will create a new directory and download the notebook. The command will look something like this:

```
jovian clone notebook-owner/notebook-id
```

4. Enter the newly created directory using `cd directory-name` and start the Jupyter notebook.

```
jupyter notebook
```

You can now access Jupyter's web interface by clicking the link that shows up on the terminal or by visiting <http://localhost:8888> on your browser. Click on the notebook file (it has a `.ipynb` extension) to open it.

Downloading the Dataset

TODO - add some explanation here

Instructions for downloading the dataset (delete this cell)

- Find an interesting dataset on this page: <https://www.kaggle.com/datasets?fileType=csv>
- The data should be in CSV format, and should contain at least 3 columns and 150 rows
- Download the dataset using the [opendatasets Python library](#).

```
!pip install jovian opendatasets --upgrade --quiet
```

Let's begin by downloading the data, and listing the files within the dataset.

```
# Change this
dataset_url = 'https://www.kaggle.com/tunguz/us-elections-dataset'
```

```
import opendatasets as od
od.download(dataset_url)
```

Kaggle dataset ID: tunguz/us-elections-dataset

0it [00:00, ?it/s]

Downloading <https://www.kaggle.com/tunguz/us-elections-dataset/download?resource=download&downloadHash=7ba6986e70e4d0e9f17ef767bfef5a5fc294114a206832fd436474da1>
to ./us-elections-dataset.zip

37806080it [00:03, 12417556.13it/s]

Extracting archive ./us-elections-dataset.zip to ./us-elections-dataset

The dataset has been downloaded and extracted.

```
# Change this
data_dir = './us-elections-dataset'
```

```
import os
os.listdir(data_dir)
```

```
['2016-precinct-house.csv',
 '1976-2018-house2.csv',
 'president-1976-2016.csv',
 '1976-2018-senate.csv',
```

```
'2016-precinct-president.csv',  
'us-2016-primary-results.csv',  
'2016-precinct-senate.csv',  
'2012_US_elect_county.csv',  
'usa-2016-presidential-election-by-county.csv']
```

Let us save and upload our work to Jovian before continuing.

```
project_name = "zerotopandas-course-project-starter" # change this (use lowercase letters)
```

```
!pip install jovian --upgrade -q
```

```
import jovian
```

```
jovian.commit(project=project_name)
```

[jovian] Attempting to save notebook..

Data Preparation and Cleaning

TODO - Write some explanation here.

Instructions (delete this cell):

- Load the dataset into a data frame using Pandas
- Explore the number of rows & columns, ranges of values etc.
- Handle missing, incorrect and invalid data
- Perform any additional steps (parsing dates, creating additional columns, merging multiple dataset etc.)

```
import jovian
```

```
jovian.commit()
```

[jovian] Attempting to save notebook..

[jovian] Updating notebook "aakashns/zerotopandas-course-project-starter" on <https://jovian.ml/>

[jovian] Uploading notebook..

[jovian] Capturing environment..

[jovian] Committed successfully! <https://jovian.ml/aakashns/zerotopandas-course-project-starter>

'<https://jovian.ml/aakashns/zerotopandas-course-project-starter>'

Exploratory Analysis and Visualization

TODO - write some explanation here.

Instructions (delete this cell)

- Compute the mean, sum, range and other interesting statistics for numeric columns
- Explore distributions of numeric columns using histograms etc.
- Explore relationship between columns using scatter plots, bar charts etc.
- Make a note of interesting insights from the exploratory analysis

Let's begin by importing matplotlib.pyplot and seaborn .

```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

TODO - Explore one or more columns by plotting a graph below, and add some explanation about it

TODO - Explore one or more columns by plotting a graph below, and add some explanation about it

TODO - Explore one or more columns by plotting a graph below, and add some explanation about it

TODO - Explore one or more columns by plotting a graph below, and add some explanation about it

TODO - Explore one or more columns by plotting a graph below, and add some explanation about it

Let us save and upload our work to Jovian before continuing

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Attempting to save notebook..
```

```
[jovian] Updating notebook "aakashns/zerotopandas-course-project-starter" on  
https://jovian.ml/
```

```
[jovian] Uploading notebook..
```

```
[jovian] Capturing environment..
```

```
[jovian] Committed successfully! https://jovian.ml/aakashns/zerotopandas-course-project-starter
```

```
'https://jovian.ml/aakashns/zerotopandas-course-project-starter'
```

Asking and Answering Questions

TODO - write some explanation here.

Instructions (delete this cell)

- Ask at least 5 interesting questions about your dataset
- Answer the questions either by computing the results using Numpy/Pandas or by plotting graphs using Matplotlib/Seaborn
- Create new columns, merge multiple dataset and perform grouping/aggregation wherever necessary
- Wherever you're using a library function from Pandas/Numpy/Matplotlib etc. explain briefly what it does

Q1: TODO - ask a question here and answer it below

Q2: TODO - ask a question here and answer it below

Q3: TODO - ask a question here and answer it below

Q4: TODO - ask a question here and answer it below

Q5: TODO - ask a question here and answer it below

Let us save and upload our work to Jovian before continuing.

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Attempting to save notebook..
```

```
[jovian] Updating notebook "aakashns/zerotopandas-course-project-starter" on  
https://jovian.ml/
```

```
[jovian] Uploading notebook..
```

```
[jovian] Capturing environment..
```

```
[jovian] Committed successfully! https://jovian.ml/aakashns/zerotopandas-course-project-starter
```

```
'https://jovian.ml/aakashns/zerotopandas-course-project-starter'
```

Inferences and Conclusion

TODO - Write some explanation here: a summary of all the inferences drawn from the analysis, and any conclusions you may have drawn by answering various questions.

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Attempting to save notebook..
```

```
[jovian] Updating notebook "aakashns/zerotopandas-course-project-starter" on  
https://jovian.ml/
```

```
[jovian] Uploading notebook..
```

```
[jovian] Capturing environment..
```

```
[jovian] Committed successfully! https://jovian.ml/aakashns/zerotopandas-course-project-starter
```

```
'https://jovian.ml/aakashns/zerotopandas-course-project-starter'
```

References and Future Work

TODO - Write some explanation here: ideas for future projects using this dataset, and links to resources you found useful.

Submission Instructions (delete this cell)

- Upload your notebook to your [Jovian.ml](https://jovian.ml) profile using `jovian.commit`.
- **Make a submission here:** <https://jovian.ml/learn/data-analysis-with-python-zero-to-pandas/assignment/course-project>
- Share your work on the forum: <https://jovian.ml/forum/t/course-project-on-exploratory-data-analysis-discuss-and-share-your-work/11684>
- Share your work on social media (Twitter, LinkedIn, Telegram etc.) and tag [@JovianML](https://twitter.com/JovianML)

(Optional) Write a blog post

- A blog post is a great way to present and showcase your work.
- Sign up on [Medium.com](https://medium.com) to write a blog post for your project.
- Copy over the explanations from your Jupyter notebook into your blog post, and [embed code cells & outputs](#)
- Check out the [Jovian.ml](https://jovian.ml) Medium publication for inspiration: <https://medium.com/jovianml>

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Attempting to save notebook..
```

```
[jovian] Updating notebook "aakashns/zerotopandas-course-project-starter" on  
https://jovian.ml/
```

```
[jovian] Uploading notebook..
```

```
[jovian] Capturing environment..
```

```
[jovian] Committed successfully! https://jovian.ml/aakashns/zerotopandas-course-project-starter
```

```
'https://jovian.ml/aakashns/zerotopandas-course-project-starter'
```