# Project Overview

**Project Overview**

In this project, you will analyze a dataset and then communicate your findings about it. You will use the Python libraries NumPy, Pandas, and Matplotlib to make your analysis easier.

Prepare for this project with: [**Intro to Data Analysis**](https://classroom.udacity.com/courses/ud170-nd)

**What do I need to install?**

You will need an installation of Python, plus the following libraries:

\* pandas

\* numpy

\* matplotlib

\* csv

We recommend installing [**Anaconda**](https://www.continuum.io/downloads), which comes with all of the necessary packages, as well as IPython notebook. You can find installation instructions [**here**](https://classroom.udacity.com/courses/ud170/lessons/5430778793/concepts/54140889150923).

**Why this Project?**

In this project, you'll go through the data analysis process and see how everything fits together. Later Nanodegree projects will focus on individual pieces of the data analysis process.

You'll use the Python libraries NumPy, Pandas, and Matplotlib, which make writing data analysis code in Python a lot easier! Not only that, these are sought-after skills by employers!

**What will I learn?**

After completing the project, you will:

* Know all the steps involved in a typical data analysis process
* Be comfortable posing questions that can be answered with a given dataset and then answering those questions
* Know how to investigate problems in a dataset and wrangle the data into a format you can use
* Have practice communicating the results of your analysis
* Be able to use vectorized operations in NumPy and Pandas to speed up your data analysis code
* Be familiar with Pandas' Series and DataFrame objects, which let you access your data more conveniently
* Know how to use Matplotlib to produce plots showing your findings

# Project Details

**How do I Complete this Project?**

This project is connected with the [**Introduction to Data Analysis**](https://classroom.udacity.com/courses/ud170) course, but depending on your background knowledge, you may not need to take the whole class to complete this project.

**Introduction**

For the final project, you will conduct your own data analysis and create a file to share that documents your findings. You should start by taking a look at your dataset and brainstorming what questions you could answer using it. Then you should use Pandas and NumPy to answer the questions you are most interested in, and create a report sharing the answers. You will not be required to use inferential statistics or machine learning to complete this project, but you should make it clear in your communications that your findings are tentative. This project is open-ended in that we are not looking for one right answer.

**Step One - Choose Your Data Set**

Click [**this link**](https://docs.google.com/document/d/e/2PACX-1vTlVmknRRnfy_4eTrjw5hYGaiQim5ctr9naaRd4V9du2B5bxpd8FEH3KtDgp8qVekw7Cj1GLk1IXdZi/pub?embedded=True) to open a document with links and information about data sets that you can investigate for this project. You **must** choose one of these datasets to complete the project.

**Step Two - Get Organized**

Eventually you’ll want to submit your project (and share it with friends, family, and employers). Get organized before you begin. We recommend creating a single folder that will eventually contain:

* The **report** communicating your findings
* Any **Python code** you wrote as part of your analysis
* The **data set** you used (which you will not need to submit)

You may wish to use Jupyter notebook, in which case you can submit both the code you wrote and the report of your findings in the same document. Otherwise, you will need to submit your report and code separately. If you would like a **notebook template** to help organize your investigation, you can find a link in the resources at the bottom of the page or you can click [**here**](https://d17h27t6h515a5.cloudfront.net/topher/2017/October/59dd0f5a_investigate-a-dataset-template/investigate-a-dataset-template.ipynb). You can also complete and submit the project in the classroom by going to the *Project Notebook* part of this lesson.

**Step Three - Analyze Your Data**

Brainstorm some questions you could answer using the data set you chose, then start answering those questions. You can find some questions in the [**data set options**](https://docs.google.com/document/d/e/2PACX-1vTlVmknRRnfy_4eTrjw5hYGaiQim5ctr9naaRd4V9du2B5bxpd8FEH3KtDgp8qVekw7Cj1GLk1IXdZi/pub?embedded=True) to help you get started.

Try and suggest questions that promote looking at relationships between multiple variables. You should aim to analyze at least one dependent variable and three independent variables in your investigation. Make sure you use NumPy and Pandas where they are appropriate!

**Step Four - Share Your Findings**

Once you have finished analyzing the data, create a report that shares the findings you found most interesting. If you use a Jupyter notebook, share your findings alongside the code you used to perform the analysis. make sure that your report text is contained in Markdown cells to clearly distinguish your comments and findings from your code work. You should also feel free to use other tools and software to craft your final report, but make sure that you can submit your report as an HTML or PDF file so that it can be opened easily.

**Step Five - Review**

Use the [**Project Rubric**](https://review.udacity.com/#!/projects/3176718735/rubric) to review your project. If you are happy with your submission, then you're ready to submit your project. If you see room for improvement, keep working to improve your project!

1. Supporting Materials

[**Investigate a Dataset - Template Notebook**](http://video.udacity-data.com.s3.amazonaws.com/topher/2017/October/59dd0f5a_investigate-a-dataset-template/investigate-a-dataset-template.ipynb)

# Project Notebook

Bajarse el Notebook

# Project Walkthrough

Ver vídeo

# Investigate a Dataset

## Project Submission

Choose one of Udacity's curated datasets and investigate it using NumPy and Pandas. Go through the entire data analysis process, starting by posing a question and finishing by sharing your findings.

## Evaluation

Use the [**Project Rubric**](https://review.udacity.com/#!/projects/3176718735/rubric) to review your project. If you are happy with your submission, then you are ready to submit! If you see room for improvement in **any** category in which you do not meet specifications, keep working!

Your project will be evaluated by a Udacity reviewer according to the same [**Project Rubric**](https://review.udacity.com/#!/projects/3176718735/rubric). Your project must "meet specifications" or "exceed specifications" in each category in order for your submission to pass.

## Submission

## What to include in your submission

1. A PDF or HTML file containing your analysis. This file should include:
   * A note specifying which dataset you analyzed
   * A statement of the question(s) you posed
   * A description of what you did to investigate those questions
   * Documentation of any data wrangling you did
   * Summary statistics and plots communicating your final results
2. Code you used to perform your analysis. If you used an Jupyter notebook, you can submit your .ipynb. Otherwise, you should submit the code separately in .py file(s).
3. A list of Web sites, books, forums, blog posts, github repositories, etc. that you referred to or used in creating your submission (add N/A if you did not use any such resources).

**Jupyter notebook instructions**

If you used Jupyter notebook to create your analysis, you can include your code directly in the notebook and do not need to create additional files to submit it separately. You will still need to export your work in a standalone format (see point 1 above). To download your notebook as an HTML file, click on File -> Download.As -> HTML (.html) within the notebook. If you get an error about "No module name", then open a terminal and try installing the missing module using pip install <module\_name> (don't include the "<" or ">" or any words following a period in the module name).

**Ready to submit your project?**

Click on the "Submit Project" button and follow the instructions to submit!

It can take us up to a week to grade the project, but in most cases it is much faster. You will get an email when your submission has been reviewed.

If you are having any problems submitting your project or wish to check on the status of your submission, please email us at [dataanalyst-project@udacity.com](mailto:dataanalyst-project@udacity.com).

What's Next?

Now that you can say you've used Python libraries on a real project, make sure to add "NumPy, Pandas, and Matplotlib" to your resume.

Don't have a resume? Choose a resume template from [**the Muse**](https://www.themuse.com/advice/the-41-best-resume-templates-ever) and start filling it out. It doesn't have to be polished - getting started and having an outline will help you later on with Udacity's resume review. After your next project, Wrangle OpenStreetMapData, we recommend optimizing your resume for an employer by completing the Resume Review in the Career: Job Search Strategies. Get a head start now!

You have not submitted the project yetSUBMIT PROJECT