|  |  |  |
| --- | --- | --- |
| **[Flights](https://www.google.com/url?q=http://stat-computing.org/dataexpo/2009/the-data.html&sa=D&source=editors&ust=1627498065197000&usg=AOvVaw3GqjDldZXnAtP3fVb36vAs)** | This dataset reports flights in the United States, including carriers, arrival and departure delays, and reasons for delays, from 1987 to 2008.   * You may want to try downloading multiple years worth of data and joining them, to do a year-by-year comparison. Not all features will be interesting for performing your exploration. * Note that the linked page points towards [another page](https://www.google.com/url?q=http://www.transtats.bts.gov/Fields.asp?Table_ID%3D236&sa=D&source=editors&ust=1627498065198000&usg=AOvVaw33oRIJNEggIpZMveh8saTb) with more detailed variable descriptions in the original, full data. | Are there certain destination or arrival cities that are home to more delays or cancellations?  What are the preferred times for flights to occur? Are there any changes over multiple years? |

**Why this project?**

Data visualization is an important skill that is used in many parts of the data analysis process. **Exploratory** data visualization generally occurs during and after the data wrangling process, and is the main method that you will use to understand the patterns and relationships present in your data. This understanding will help you approach any statistical analyses and will help you build conclusions and findings. This process might also illuminate additional data cleaning tasks to be performed. **Explanatory** data visualization techniques are used after generating your findings, and are used to help communicate your results to others. Understanding design considerations will make sure that your message is clear and effective. In addition to being a good producer of visualizations, going through this project will also help you be a good consumer of visualizations that are presented to you by others.

**What will I learn?**

After completing this project, you will be able to:

* Supplement statistics with visualizations to build understanding of data.
* Choose appropriate plots, limits, transformations, and aesthetics to explore a dataset, allowing you to understand distributions of variables and relationships between features.
* Use design principles to create effective visualizations for communicating findings to an audience.

If you’re finding your own dataset…

Your dataset should:

* include at least 600 observations. (This is the number of rows after tidying your data - see the bullet point below about tidy data.)
* include at least eight variables.
* include at least one qualitative / categorical variable. (This can also be engineered / created.)
* include at least one numeric variable.
* be in a tidy format. In a nutshell, tidy data has each row as a single observation and each column reporting a single variable. You can read more about tidy data in Hadley Wickham’s paper [[here](https://www.google.com/url?q=http://vita.had.co.nz/papers/tidy-data.pdf&sa=D&source=editors&ust=1627498065210000&usg=AOvVaw3m3xh1GYYgX0oETTrdXY72)]. You may need to do some cleaning and reshaping to tidy your dataset, before you actually get started with your exploration.
* be in a common data format. This includes .csv, .tsv, .txt, and .xls. Basically, there should be a reasonable pandas.read\_\*() function to open up your data in a tidy format as a pandas DataFrame.

This project is divided into two major parts. In the first part, you will conduct an **exploratory** data analysis on a dataset of your choosing. You will use Python data science and data visualization libraries to explore the dataset’s variables and understand the data’s structure, oddities, patterns and relationships. The analysis in this part should be structured, going from simple univariate relationships up through multivariate relationships, but it does not need to be clean or perfect. There is no one single answer that needs to come out of a given dataset. This part of the project is your opportunity to ask questions of the data and make your own discoveries. It’s important to keep in mind that sometimes exploration can lead to dead ends, and that it can take multiple steps to dig down to what you’re truly looking for. Be patient with your steps, document your work carefully, and be thorough in the perspective that you choose to take with your dataset.

In the second part, you will take your main findings from your exploration and convey them to others through an **explanatory** analysis. To this end, you will create a slide deck that leverages polished, explanatory visualizations to communicate your results. This part of the project should make heavy use of the first part of the project. Select one or two major paths in your exploration, choose relevant visualizations along that path, and then polish them to construct a story for your readers to understand what you found.

**Step 1.1: Choose your Dataset**

First, you will choose a dataset from the Dataset Options.

Download the Dataset Options file for full details & descriptions from the Resources Tab.

1. Click on Resources in the leftmost panel of your classroom
2. Click the File Name to start download

**Quick List Below**:

Dataset Options

Ford GoBike System Data - Available at the bottom of the page

[Flights](https://www.google.com/url?q=http://stat-computing.org/dataexpo/2009/the-data.html&sa=D&ust=1554486256017000)

[Loan Data from Prosper](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv&sa=D&ust=1554486256021000) with [Prosper Data Dictionary to Explain Dataset's Variables](https://www.google.com/url?q=https://docs.google.com/spreadsheet/ccc?key%3D0AllIqIyvWZdadDd5NTlqZ1pBMHlsUjdrOTZHaVBuSlE%26usp%3Dsharing&sa=D&ust=1554486256024000)

[PISA Data](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud507/pisa2012.csv.zip&sa=D&ust=1554482573645000) with [PISA Data Dictionary to Explain Dataset's Variables](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud507/pisadict2012.csv&sa=D&ust=1554482573645000)

Or select your own dataset! See guidelines in the Dataset Options download in the Resources tab on whether or not a dataset will be appropriate for use in this project Remember that finding and cleaning your own data set could take significant time and effort!

A Google Doc download option with identical info is available below as well, if you prefer it. This is not accessible on all networks. [Google Doc Download](https://docs.google.com/document/d/e/2PACX-1vQmkX4iOT6Rcrin42vslquX2_wQCjIa_hbwD0xmxrERPSOJYDtpNc_3wwK_p9_KpOsfA6QVyEHdxxq7/pub?embedded=True)

**Step 1.2: Explore Your Data**

It’s time to get to the interesting bits. Explore your data and document your findings in a report. The report should briefly introduce the dataset, then systematically walk through the points of exploration that you conducted. You should have headers and text that organize your thoughts and findings. Visualizations in this part of the project need not be completely polished: this is just your own exploration at this point. However, you should still make sure that you adhere to principles of using appropriate plot types and encodings so that accurate conclusions can be drawn, and that you have enough comments and labeling so that when you return to your work, you can quickly grasp your analysis steps.

If you use a Jupyter Notebook for this step of the project, don’t forget to export the notebook as an html file for the project submission.

**Step 2.1: Document your Story**

At the end of your exploration, you probably have a bunch of things that you’ve discovered. Now it’s time to organize your findings and select a story that you will convey to others. In your readme document, you should summarize your main findings and reflect on the steps you took in your data exploration. You should also lay out the key insights that you want to convey in your explanatory report as well as any changes to visualizations, or note new visualizations that will be created to bridge between your insights.

**Step 2.2: Create your Slide Deck**

Follow the plans you laid out in the previous step and create a slide deck with explanatory data visualizations to tell a story about the data you explored. You can start with code that you used in your exploration, but you should make sure that the code is revised so that your plots are polished. Make sure that you also pay attention to aspects of design integrity in your revisions.

**Step 2.3: (Optional) Get Feedback**

Though not required, it is highly recommended that you try to get feedback from at least one person before you submit your project. By sharing your work with others, you can get input from a different perspective that catches things that you may have originally missed. Share your slide deck with someone in person and have them provide live feedback on what they get from your slide deck. Alternatively, you can also share your work with your fellow students. Post a message in a student community channel for this project with a link to your project and ask for feedback. Be sure to keep an eye out for others who are also seeking feedback and return the favor!

You might need to ask specific questions to prompt your reader. The following questions might be good starters; be sure to follow up or come up with your own questions:

* What do you notice about each visualization?
* What questions do you have about the data?
* What relationships do you notice?
* What do you think is the main takeaway from the slide deck?
* Is there anything that you don’t understand from the plots?

If you get feedback from others, then add their feedback to your readme document. Note what changes you make to your slide deck and designs based on that feedback. You can also include feedback from your reviewer as part of this revision process.

**Step 2.4: Review and Submit the Project**

There’s one last thing to do before you submit your project. You should closely read the [project rubric](https://review.udacity.com/#!/projects/8ff9475b-3d6b-4c5b-9593-96794db62987/rubric), which your reviewer will use to evaluate your work. Perform a self-review to assess the quality of your work. If there is any rubric point on which you don’t feel confident, then go back and make improvements before submitting your project for review. In fact, it’s a good idea to *look at the*[*project rubric*](https://review.udacity.com/#!/projects/8ff9475b-3d6b-4c5b-9593-96794db62987/rubric)*now*, before you get started. This way, you have a clear idea of what points to aim for as you go through the project steps.

**Supporting Materials**

* [Dataset for Communicate Data Findings Project](https://video.udacity-data.com/topher/2019/April/5ca78b26_dataset-project-communicate-data-findings/dataset-project-communicate-data-findings.pdf)
* [201902-Fordgobike-Tripdata](https://video.udacity-data.com/topher/2020/October/5f91cf38_201902-fordgobike-tripdata/201902-fordgobike-tripdata.csv)

To help you get rolling with the project, we’ve provided two .zip folders for you to download, by clicking on the headers below, or by checking out the Resources tab.

[**Project Template**](https://s3.amazonaws.com/video.udacity-data.com/topher/2018/April/5ac6ccef_communicate-data-project-template/communicate-data-project-template.zip)

This archive contains four files that will help you get organized with your project.

* **readme.md** - This Markdown file contains sections that you should fill out as you select your dataset, complete your exploration, and plan your explanatory analysis. You can open up markdown files using any plain text editor: the format is meant to allow for good readability both as plain text and rendered as HTML. If you need a primer for Markdown syntax, you can find the documentation on the original Markdown specifications [here](https://daringfireball.net/projects/markdown/syntax).
* **exploration\_template.ipynb** - This Jupyter Notebook contains section templates to help you organize your exploration, starting from loading in the data, working through univariate visualizations, and ending with bivariate and multivariate exploration. At the end of each section, there are questions to help you summarize your findings.
* **slide\_deck\_template.ipynb** - This Jupyter Notebook contains starter cells to help you organize your slide deck deliverable. These cells provide an example of how the slide deck should be organized, including pre-set slideshow settings.
* **output\_toggle.tpl** - This template file can be used with nbconvert to export your slide deck. This adds extra functionality to the slide deck by hiding the code to start, only making it visible if the reader clicks on the output (which should mostly be visualizations in the case of this project). This template file was taken from [this page](https://github.com/damianavila/blog/blob/master/posts/hide-the-input-cells-from-your-ipython-slides.ipynb), written by one of the contributors to the nbconvert project.

The Markdown and Jupyter notebook files have quoted sections with tips on how to use the templates. If you make use of these templates, make sure that you remove these quoted sections before you export and submit your project!

[**Example Project**](https://s3.amazonaws.com/video.udacity-data.com/topher/2018/August/5b7de78c_communicate-data-example-project/communicate-data-example-project.zip)

This archive contains seven files that serve as an example of what a final project submission could look like. This example project was built using the project template files, and using the diamond price dataset that was incorporated in the course associated with this project. You cannot use this diamonds dataset to complete your project. It is recommended that you take some time to look over the example project to get a sense for what you will need to do in your own analysis. To view the slide deck, you will need to use the expression (all one line):

*jupyter nbconvert Example\_Project\_Diamonds\_Part2.ipynb --to slides --post serve --template output\_toggle*

from the terminal or command line, rather than just opening up the associated html file directly.

2009

Data Expo 2009 - Airline on-time performance

Have you ever been stuck in an airport because your flight was delayed or cancelled and wondered if you could have predicted it if you'd had more data? This is your chance to find out.

The results

We had a total of [nine entries](http://stat-computing.org/dataexpo/2009/posters), and turn out at the poster session at the JSM was great, with plenty of people stopping by to find out why their flights were delayed.

The data

The data set is available for download [here](http://ww2.amstat.org/sections/graphics/datasets/DataExpo2009.zip).  
[The data](http://stat-computing.org/dataexpo/2009/the-data.html) consists of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008. This is a large dataset: there are nearly 120 million records in total, and takes up 1.6 gigabytes of space compressed and 12 gigabytes when uncompressed. To make sure that you're not overwhelmed by the size of the data, we've provide two brief introductions to some useful tools: [linux command line tools](http://stat-computing.org/dataexpo/2009/unix-tools.html) and [sqlite](http://stat-computing.org/dataexpo/2009/sqlite.html), a simple sql database.

The challenge

The aim of the data expo is to provide a **graphical** summary of important features of the data set. This is intentionally vague in order to allow different entries to focus on different aspects of the data, but here are a few ideas to get you started:

* When is the best time of day/day of week/time of year to fly to minimise delays?
* Do older planes suffer more delays?
* How does the number of people flying between different locations change over time?
* How well does weather predict plane delays?
* Can you detect cascading failures as delays in one airport create delays in others? Are there critical links in the system?

You are also welcome to work with interesting subsets: you might want to compare flight patterns before and after 9/11, or between the pair of cities that you fly between most often, or all flights to and from a major airport like Chicago (ORD). Smaller subsets may also help you to match up the data to [other interesting datasets](http://stat-computing.org/dataexpo/2009/supplemental-data.html).

<https://community.amstat.org/jointscsg-section/dataexpo/dataexpo2009>

This project is divided into two major parts. In the first part, you will conduct an **exploratory** data analysis on a dataset of your choosing. You will use Python data science and data visualization libraries to explore the dataset’s variables and understand the data’s structure, oddities, patterns and relationships. The analysis in this part should be structured, going from simple univariate relationships up through multivariate relationships, but it does not need to be clean or perfect. There is no one single answer that needs to come out of a given dataset. This part of the project is your opportunity to ask questions of the data and make your own discoveries. It’s important to keep in mind that sometimes exploration can lead to dead ends, and that it can take multiple steps to dig down to what you’re truly looking for. Be patient with your steps, document your work carefully, and be thorough in the perspective that you choose to take with your dataset.

In the second part, you will take your main findings from your exploration and convey them to others through an **explanatory** analysis. To this end, you will create a slide deck that leverages polished, explanatory visualizations to communicate your results. This part of the project should make heavy use of the first part of the project. Select one or two major paths in your exploration, choose relevant visualizations along that path, and then polish them to construct a story for your readers to understand what you found.

## Step 1.1: Choose your Dataset

First, you will choose a dataset from the Dataset Options.

Download the Dataset Options file for full details & descriptions from the Resources Tab.

1. Click on Resources in the leftmost panel of your classroom
2. Click the File Name to start download

**Quick List Below**:

Dataset Options

Ford GoBike System Data - Available at the bottom of the page

[**Flights**](https://www.google.com/url?q=http://stat-computing.org/dataexpo/2009/the-data.html&sa=D&ust=1554486256017000)

[**Loan Data from Prosper**](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv&sa=D&ust=1554486256021000) with [**Prosper Data Dictionary to Explain Dataset's Variables**](https://www.google.com/url?q=https://docs.google.com/spreadsheet/ccc?key%3D0AllIqIyvWZdadDd5NTlqZ1pBMHlsUjdrOTZHaVBuSlE%26usp%3Dsharing&sa=D&ust=1554486256024000)

[**PISA Data**](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud507/pisa2012.csv.zip&sa=D&ust=1554482573645000) with [**PISA Data Dictionary to Explain Dataset's Variables**](https://www.google.com/url?q=https://s3.amazonaws.com/udacity-hosted-downloads/ud507/pisadict2012.csv&sa=D&ust=1554482573645000)

Or select your own dataset! See guidelines in the Dataset Options download in the Resources tab on whether or not a dataset will be appropriate for use in this project Remember that finding and cleaning your own data set could take significant time and effort!

A Google Doc download option with identical info is available below as well, if you prefer it. This is not accessible on all networks. [**Google Doc Download**](https://docs.google.com/document/d/e/2PACX-1vQmkX4iOT6Rcrin42vslquX2_wQCjIa_hbwD0xmxrERPSOJYDtpNc_3wwK_p9_KpOsfA6QVyEHdxxq7/pub?embedded=True)

## Step 1.2: Explore Your Data

It’s time to get to the interesting bits. Explore your data and document your findings in a report. The report should briefly introduce the dataset, then systematically walk through the points of exploration that you conducted. You should have headers and text that organize your thoughts and findings. Visualizations in this part of the project need not be completely polished: this is just your own exploration at this point. However, you should still make sure that you adhere to principles of using appropriate plot types and encodings so that accurate conclusions can be drawn, and that you have enough comments and labeling so that when you return to your work, you can quickly grasp your analysis steps.

If you use a Jupyter Notebook for this step of the project, don’t forget to export the notebook as an html file for the project submission.

## Step 2.1: Document your Story

At the end of your exploration, you probably have a bunch of things that you’ve discovered. Now it’s time to organize your findings and select a story that you will convey to others. In your readme document, you should summarize your main findings and reflect on the steps you took in your data exploration. You should also lay out the key insights that you want to convey in your explanatory report as well as any changes to visualizations, or note new visualizations that will be created to bridge between your insights.

## Step 2.2: Create your Slide Deck

Follow the plans you laid out in the previous step and create a slide deck with explanatory data visualizations to tell a story about the data you explored. You can start with code that you used in your exploration, but you should make sure that the code is revised so that your plots are polished. Make sure that you also pay attention to aspects of design integrity in your revisions.

## Step 2.3: (Optional) Get Feedback

Though not required, it is highly recommended that you try to get feedback from at least one person before you submit your project. By sharing your work with others, you can get input from a different perspective that catches things that you may have originally missed. Share your slide deck with someone in person and have them provide live feedback on what they get from your slide deck. Alternatively, you can also share your work with your fellow students. Post a message in a student community channel for this project with a link to your project and ask for feedback. Be sure to keep an eye out for others who are also seeking feedback and return the favor!

You might need to ask specific questions to prompt your reader. The following questions might be good starters; be sure to follow up or come up with your own questions:

* What do you notice about each visualization?
* What questions do you have about the data?
* What relationships do you notice?
* What do you think is the main takeaway from the slide deck?
* Is there anything that you don’t understand from the plots?

If you get feedback from others, then add their feedback to your readme document. Note what changes you make to your slide deck and designs based on that feedback. You can also include feedback from your reviewer as part of this revision process.

## Step 2.4: Review and Submit the Project

There’s one last thing to do before you submit your project. You should closely read the [**project rubric**](https://review.udacity.com/#!/projects/8ff9475b-3d6b-4c5b-9593-96794db62987/rubric), which your reviewer will use to evaluate your work. Perform a self-review to assess the quality of your work. If there is any rubric point on which you don’t feel confident, then go back and make improvements before submitting your project for review. In fact, it’s a good idea to look at the [***project rubric***](https://review.udacity.com/#!/projects/8ff9475b-3d6b-4c5b-9593-96794db62987/rubric) now, before you get started. This way, you have a clear idea of what points to aim for as you go through the project steps.

#### Supporting Materials

* [**Dataset for Communicate Data Findings Project**](https://video.udacity-data.com/topher/2019/April/5ca78b26_dataset-project-communicate-data-findings/dataset-project-communicate-data-findings.pdf)
* [**201902-Fordgobike-Tripdata**](https://video.udacity-data.com/topher/2020/October/5f91cf38_201902-fordgobike-tripdata/201902-fordgobike-tripdata.csv)

NEXT

**PROJECT SPECIFICATION**

**Communicate Data Findings**

Code Quality

| CRITERIA | MEETS SPECIFICATIONS |
| --- | --- |
| Does the code work? | All code is functional (i.e. no errors are thrown by the code). Warnings are okay, as long as they are not a result of poor coding practices. |
| Does the project follow good coding practices? | The project uses functions and loops where possible to reduce repetitive code. Comments and docstrings are used as needed to document code functionality. |

Exploratory Data Analysis

| CRITERIA | MEETS SPECIFICATIONS |
| --- | --- |
| Is the data explored systematically? | The project appropriately uses univariate, bivariate, and multivariate plots to explore many relationships in the data set. Reasoning is used to justify the flow of the exploration. |
| Are questions and observations documented in the report? | Questions and observations are placed regularly throughout the report, after each plot or set of related plots. |
| Is the data visualized using appropriate plot types, encodings, and parameter choices? | Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted. This includes choice of appropriate plot type, data encodings, transformations, and labels as needed. |

Explanatory Data Analysis

| CRITERIA | MEETS SPECIFICATIONS |
| --- | --- |
| Have the main findings from the exploration been documented? | A section in the submitted materials includes a summary of main findings that reflects on the steps taken during the data exploration. The section also describes the key insights that are conveyed by the explanatory presentation. |
| Does the presentation clearly convey key insights? | A slideshow is provided, with at least three visualizations used in the presentation to convey key insights. These key insights match those documented in the summary. Each visualization is associated with comments that accurately depict their purpose. |
| Are the plots polished? | All plots in the presentation have an appropriate title with labeled axes and legends. Labels include units as needed. Plot type, encodings, and transformations are all appropriate. |

**Suggestions to Make Your Project Stand Out!**

* During the exploration, use a variety of plot types to explore different relationships in the dataset. Be willing to investigate unexpected relationships and don’t be afraid of finding a dead-end in your exploration.
* As part of your exploration, document your thought processes to justify the steps you take.
* When you select key insights for your explanatory presentation, focus on one or two paths that tell a compelling story.
* When planning your explanation’s flow, document design decisions that make your visualizations information-rich but still easy to read.
* Gather feedback from others to get a different perspective on your explanatory presentation. Document that feedback and note any changes in your designs based on that feedback.

To help you get rolling with the project, we’ve provided two .zip folders for you to download, by clicking on the headers below, or by checking out the Resources tab.

[**Project Template**](https://s3.amazonaws.com/video.udacity-data.com/topher/2018/April/5ac6ccef_communicate-data-project-template/communicate-data-project-template.zip)

This archive contains four files that will help you get organized with your project.

* **readme.md** - This Markdown file contains sections that you should fill out as you select your dataset, complete your exploration, and plan your explanatory analysis. You can open up markdown files using any plain text editor: the format is meant to allow for good readability both as plain text and rendered as HTML. If you need a primer for Markdown syntax, you can find the documentation on the original Markdown specifications [**here**](https://daringfireball.net/projects/markdown/syntax).
* **exploration\_template.ipynb** - This Jupyter Notebook contains section templates to help you organize your exploration, starting from loading in the data, working through univariate visualizations, and ending with bivariate and multivariate exploration. At the end of each section, there are questions to help you summarize your findings.
* **slide\_deck\_template.ipynb** - This Jupyter Notebook contains starter cells to help you organize your slide deck deliverable. These cells provide an example of how the slide deck should be organized, including pre-set slideshow settings.
* **output\_toggle.tpl** - This template file can be used with nbconvert to export your slide deck. This adds extra functionality to the slide deck by hiding the code to start, only making it visible if the reader clicks on the output (which should mostly be visualizations in the case of this project). This template file was taken from [**this page**](https://github.com/damianavila/blog/blob/master/posts/hide-the-input-cells-from-your-ipython-slides.ipynb), written by one of the contributors to the nbconvert project.

The Markdown and Jupyter notebook files have quoted sections with tips on how to use the templates. If you make use of these templates, make sure that you remove these quoted sections before you export and submit your project!

[**Example Project**](https://s3.amazonaws.com/video.udacity-data.com/topher/2018/August/5b7de78c_communicate-data-example-project/communicate-data-example-project.zip)

This archive contains seven files that serve as an example of what a final project submission could look like. This example project was built using the project template files, and using the diamond price dataset that was incorporated in the course associated with this project. You cannot use this diamonds dataset to complete your project. It is recommended that you take some time to look over the example project to get a sense for what you will need to do in your own analysis. To view the slide deck, you will need to use the expression (all one line):

jupyter nbconvert Example\_Project\_Diamonds\_Part2.ipynb --to slides --post serve --template output\_toggle

from the terminal or command line, rather than just opening up the associated html file directly.

Project Submission

Use Python visualization libraries, including matplotlib and seaborn, to systematically explore a selected dataset, starting from plots of single variables and building up to plots of multiple variables. Then, produce a short presentation that illustrates interesting properties, trends, and relationships that you discovered in your selected dataset.

**Evaluation**

Use the instructions in the project lesson to complete the steps in the project. Once you're done, use the [**Project Rubric**](https://review.udacity.com/#!/rubrics/1795/view) to review your project. If you see room for improvement in **any** category in which you do not meet specifications, keep working! If you are happy with your project, then you are ready to submit it!

Remember, your project will be evaluated by a Udacity reviewer according to the same [**Project Rubric**](https://review.udacity.com/#!/rubrics/1795/view). Your project must "Meet Specifications" in each category in order for your submission to pass.

**Submission**

If you're ready to submit your project, make sure that you collect the following files in a .zip file:

1. A report with your **exploratory** data analysis, in PDF or HTML format. If you used a Jupyter Notebook to conduct your analysis, you should also include the original .ipynb file in your submission. If you did not, you should include the code you used in your exploration as .py scripts.
2. A slide deck presentation with your **explanatory** analysis, in PDF or HTML format. If you used a Jupyter Notebook, include the original .ipynb file with your submission and any template file used to render the slide deck.
3. A **readme** document, in plain text, Markdown, or PDF format, including the following information:
   * Which dataset you chose. If not part of Udacity's dataset options, document the source of your data.
   * Main findings from the exploratory data analysis, and how you chose the results to put in your explanatory analysis.
   * If you obtained feedback from others for your explanatory designs, document them here.
   * List of resources used during the creation of the project. This includes web sites, books, forums, blog posts, and GitHub repositories.
4. If you chose a dataset that is **not** in the [**Dataset Options**](https://docs.google.com/document/d/e/2PACX-1vQmkX4iOT6Rcrin42vslquX2_wQCjIa_hbwD0xmxrERPSOJYDtpNc_3wwK_p9_KpOsfA6QVyEHdxxq7/pub?embedded=True) document, include the **dataset** used to perform the analyses. If the dataset is too large, then make sure the readme documents where the data can be found so that the reviewer can check your work as needed.

**Ready to submit your project?**

Click on the "Submit Project" button below 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 [**review-support@udacity.com**](mailto:review-support@udacity.com).