

**CAREER***FOUNDRY*

# **Achievement 6 Project Brief: Advanced Analytics & Dashboard Design**

# Objective

To build an **interactive dashboard** visually **showcasing well-curated results** of an **advanced exploratory analysis conducted in Python**.

## Context

For this Achievement, you'll **select your own data to analyze**. You'll first conduct an **exploratory visual analysis in Python** before identifying connections between variables worth further exploration. Then, after **developing hypotheses**, you'll use various **advanced analytical approaches** to help **test them** before **presenting your results in a Tableau dashboard/storyboard**.

Data dashboards are an effective tool for presenting data in an accessible and tangible way. **Your dashboard will tell the story of your analytical journey and should contain a *curation* of the key results you discovered throughout this Achievement. Not all of your results will fit into the dashboard—make sure to include any additional analyses you conducted in a GitHub repository.**

## Data Requirements

Your chosen data must meet specific criteria so that you can conduct the procedures explored in this Achievement (and develop the necessary skills for a junior analyst). You'll likely **need to source multiple data sets while working through the Achievement**; however, we advise you to **start with one main data set**. Keep in mind that data sourcing can be very time-consuming!

The data set(s) you choose must:

- Be **open source**;
- Come from an **authentic/authoritative source**;
- Include **non-anonymized column names**;
- Be **no more than 3 years old** (up to a **maximum of 10 years** if you've found a perfect data set for your needs and no newer data is available);
- Contain **at least 2 continuous variables** (excluding index or ID variables, dates, years, etc.);
- Contain **at least 2 categorical variables** (excluding index or ID variables, dates, years, etc.);
- Contain **at least 1,500 rows**;
- Include **a geographical component with at least 2 different values** (e.g., countries, continents, U.S. states, cities, latitude and longitude values—**anything you can visualize on a map!**).

**Note:** Later in the Achievement, you'll source a time series data set, too. You'll learn more about that in the corresponding Exercise, so don't worry about it for now.

In summary, the **most important feature** your data set(s) must contain is a **geographic object**—you'll need this data to **conduct a geospatial analysis**. In Exercise 6.1, we've suggested different data types you could use for this project based on your professional experience and goals.

If you're struggling to find a viable data set of your own, we've sourced some for you that meet the above criteria:

- [Boat Sales](#)
- [New York Citi Bikes](#)
- [World Happiness Report 2015–2019](#)
- [Airbnb Amsterdam](#)
- [Brazilian E-Commerce](#)
- [House Sales in King County, USA](#)
- [Zomato Bangalore Restaurants](#) (to be used with this [JSON](#))
- [Medical Cost Personal Datasets](#) (to be used with this [JSON](#))
- [World University Rankings](#)
- [Chocolate Bar Ratings](#)
- [Gun Violence Data](#)
- [UFC-Fight historical data from 1993 to 2021](#)

## Analysis Criteria

You'll conduct the following analyses throughout the Achievement (note that you won't need to include all results in your final dashboard, just those that help tell your story):

- Exploratory analysis through visualizations (scatterplots, correlation heatmaps, pair plots, and categorical plots);
- Geospatial analysis using a shapefile;
- Regression analysis;
- Cluster analysis;
- Time-series analysis;
- Analysis narrative and final results (presented in your dashboard).

# Dashboard Requirements

Your dashboard needs to tell the story of your analytical journey, so it should contain a curation of the key results you discovered throughout the Achievement. Based on your findings, you can decide which visuals and procedures are helpful to include.

## Your final dashboard must:

- Be designed with a use case in mind (answering key guidance questions);
- Be created in Tableau Public;
- Be interactive;
- Adhere to visual design best practices;
- Include an introduction page that describes the project (data and purpose);
- Include relevant result(s) of initial visual exploratory analysis;
- Include an explanation for how the results of the exploratory analysis resulted in defining research questions and/or hypotheses;
- Contain a geospatial component;
- Address the defined questions/hypotheses using advanced analytical techniques. For example,
  - Regression analysis;
  - Cluster analysis;
  - Time-series analysis.
- Include a results summary page explaining how the results do (or don't!) address your initial research questions/hypotheses;
- Include details on the limitations of the project;
- Include a proposal of the next steps for further analysis.

**Note:** You can include any analyses that don't make it onto your final dashboard in your GitHub repository.

# GitHub Repository Requirements

You'll create your dashboard in Tableau, but any analysis you conducted in Python should also be available to viewers. Using a GitHub repository will give you a place to include any steps that don't fit into your dashboard's narrative.

Your GitHub repository must include:

- Your Python code;
- A logical folder structure;
- Folders and files that follow industry-standard naming conventions;
- Portfolio-ready Jupyter scripts for every task in the Achievement (complete with code comments, organized structure, and clean, functioning code);
- A README file containing:
  - A description of the project;
  - Details of the data source(s);
  - Research questions;
  - Cleaning procedures (from Exercise 6.1);
  - A link to your Tableau dashboard.

# Your Project Deliverables

Throughout this Achievement, you'll work from Exercise to Exercise, **submitting a deliverable in each task that directly contributes to the final product**—your **data dashboard**. You'll begin by sourcing your data, exploring it, and defining the scope of your project. Then, you'll conduct your analyses before finally building your data dashboard.

Here's a breakdown of your project deliverables by Exercise.

## Exercise 6.1: Sourcing Open Data

- **Source the data for your project**, ensuring it meets the requirements outlined in this brief.
- **Prepare your data for analysis (e.g., cleaning)**.
- **Define questions** you'd like to explore based on your understanding of the data.
- Create a **document outlining the initial steps conducted**.

## Exercise 6.2: Exploring Relationships

- Guided by the questions you defined in the previous task, **conduct exploratory visual analysis using relevant Python libraries**.
- **Define hypotheses to test**.

## Exercise 6.3: Geographical Visualizations with Python

- **Source a shapefile containing the location data in your main project data set**.
- **Wrangle, clean, and merge** data files in preparation for analysis.
- Create a **choropleth map using relevant Python libraries to conduct geospatial analysis**.

## Exercise 6.4: Supervised Machine Learning: Regression

- **State your hypothesis**.
- Select the **relevant variables**.
- Prepare your data for a **regression analysis**.
- **Split the data into two sets: training and test sets**.
- **Run a linear regression** on the data and **analyze the model performance statistics**.

## Exercise 6.5: Unsupervised Machine Learning: Clustering

- Prepare your data for a cluster analysis.
- Use the elbow technique to determine the optimal number of clusters.
- Run the k-means algorithm.
- Attach a new column to your dataframe with the resulting clusters.
- Create a variety of different visualizations using your clustered data.
- Calculate the descriptive statistics for your clusters using the groupby() function and discuss your findings and any proposed next steps.

## Exercise 6.6: Sourcing & Analyzing Time Series Data

- Source time-series data relevant to your project data via an API.
- Subset your data if necessary so that it contains only relevant historical data.
- Visualize the data in a line plot and decompose its structure.
- Conduct a Dickey-Fuller test and plot autocorrelations to test for stationarity.
- Perform differencing to stationarize non-stationary data.

## Exercise 6.7: Creating Data Dashboards

- Define the use case for your dashboard.
- Outline your dashboard's contents based on the curated results of your analyses.
- Create a dashboard/storyboard in Tableau that meets the requirements of this brief.
- Publish your storyboard to Tableau Public.
- Create a portfolio-ready GitHub repository for your project that meets the requirements of this brief.