For Project 1, you will work with your group to find and analyze a dataset of your choice.

For this project, you can focus your efforts within a specific industry, as detailed in the following examples.

**Finance**

Exploratory data analysis is used by many individuals within the finance industry, including investment banking professionals, private equity analysts, lending analysts, financial administrators, and real estate professionals.

Exploratory data analysis is used for the following tasks in the financial sector:

* Identifying deals
* Analyzing private equity markets
* Researching arbitrage opportunities
* Evaluating liquidity
* Keeping up to date with finance and refinance trends

**Project Examples**

* Equity Trading: While working for a large equity-trading company, you’re tasked with researching a client’s portfolio. Your client wants to invest in telecom stocks and needs expert analysis to make the right decision. Using the [Nasdaq Data APILinks to an external site.](https://data.nasdaq.com/tools/api), pull a year’s worth of trading data for the major cell phone providers: AT&T, T-Mobile, and Verizon. Which stocks are trending upward? Which are trending downward? Based on the data, what would you recommend to your client?
* New-Car Loan Analysis: People have been financing higher car values over longer amounts of time. Explore what is driving this trend. Search for answers by using data collected from the [Federal Reserve Economic Data (FRED)Links to an external site.](https://fred.stlouisfed.org/series/DTCTLVENANQ). What other questions can you answer with the data? What do your results suggest about the time value of money? What about the impact of these loans as time goes on?

**Healthcare**

Exploratory data analysis is used by many individuals within the healthcare field, including clinical data analysts, pharmaceutical testers, healthcare-economics researchers, senior policy analysts, compliance operations analysts, and public health informatics scientists.

Exploratory data analysis is important for understanding the following healthcare considerations:

* Predicting and diagnosing illnesses
* Improving patient safety
* Reducing time to diagnosis
* Increasing our understanding of disease risks and causes
* Developing stronger prevention strategies

**Project Examples**

* Mental Health in Tech: People working in tech are often at their desks for extended amounts of time. Explore how this trend correlates with mental health. Examine the [data collected through surveysLinks to an external site.](https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey) and search for trends. Find out if there is a link between mental health and companies that offer wellness programs. What do the results show you about the state of mental health in tech? Can you suggest steps that companies can take to help their employees?
* Personal Fitness Analyst: Research whether working out helps a person become more active overall. Use data collected by the [Samsung Health applicationLinks to an external site.](https://www.kaggle.com/datasets/aroojanwarkhan/fitness-data-trends) to uncover relevant trends. What do the results tell you about individuals using this app? Have their lifestyles become more active? Less? Remained the same?

**Custom**

We’ve only specified healthcare and finance, but any industry can benefit from exploratory data analysis.

The following professionals also use data and can benefit from exploratory data analysis:

* Natural and environmental scientists
* Marketing professionals
* Information security analysts
* Business intelligence analysts

**Project Examples**

* Private Investigator: Use [aggregate crime dataLinks to an external site.](https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i) from different police precincts in a city to uncover criminal activity patterns. Consider that [most crime in New York City takes place in the summerLinks to an external site.](https://www.nydailynews.com/new-york/nyc-crime/daily-news-analysis-reveals-crime-rankings-city-subway-system-article-1.1836918). Find out if you are able to uncover similar patterns in your city. What do your results suggest about how police should plan their patrols? What do your results suggest about how law enforcement resources should be distributed over the calendar year?
* Uber Rides and Weather: No one likes to walk in subzero temperatures or scorching heat. Do people use Uber more when the weather is uncomfortable? Using [Uber ride data from KaggleLinks to an external site.](https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city) and data from a weather API, find out if people take Uber more during summer and winter months, and if there are relationships between daily temperature and ride frequency. What do the results tell you about surge-pricing strategies and commuter habits?

**Working with Your Group**

When working on an online group project, it’s crucial to meet with your group and communicate regularly. Plan for significant collaboration time outside of class. The following tips can help you make the most of your time:

* Decide how you’re going to communicate with your group members when you begin. Create a Slack channel, exchange phone numbers, and ensure that the group knows each group member’s available working hours.
* Set up an agile project by using [GitHub ProjectsLinks to an external site.](https://docs.github.com/en/free-pro-team@latest/github/managing-your-work-on-github/managing-project-boards) so that your group can track tasks.
* Create internal milestones to ensure that your group is on track. Set due dates for these milestones so that you have a timeline for completing the project. Some of these milestones might include:
  + Project ideation
  + Data fetching/API integration
  + Data analysis
  + Testing
  + Creating documentation
  + Creating the presentation

Since this is a two-week project, make sure that you have done at least half of your project by the end of the first week in order to stay on track.

Although you will divide the work among the group members, it’s essential to collaborate and communicate while working on different parts of the project. Be sure to check in with your teammates regularly and offer support.

**Support and Resources**

Your instructional team will provide support during classes and office hours. You will also have access to learning assistants and tutors to help you with topics as needed. Make sure to take advantage of these resources as you collaborate with your group on this first project.

**Requirements**

**Completed Analysis Uploaded to GitHub (20 points)**

* Final data analysis contains ample and complete information in README file (10 points)
* Final repository is acceptable for professional quality presentation (10 points)

**Visualizations (20 points)**

* 6–8 visualizations of data (at least two per question) (10 points)
* Clear and accurate labeling of images (5 points)
* Visualizations supported with ample and precise explanation (5 points)

**Analysis and Conclusion (20 points)**

* Write-up summarizes major findings and implications at a professional level (5 points)
* Each question in the project proposal is answered with precise descriptions and findings (5 points)
* Findings are strongly supported with numbers and visualizations (5 points)
* Each question response is supported with a well-discerned statistical analysis from lessons (e.g., aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis) (5 points)

**Group Presentation (20 points)**

* All group members spoke during the presentation (5 points)
* Group was well prepared (5 points)
* Presentation is relevant to material (5 points)
* Presentation maintains audience interest (5 points)

**Slide Deck (20 points)**

* Slides are visually clean and professional (5 points)
* Slides are relevant to material (5 points)
* Slides effectively demonstrate the project (5 points)
* Slides are clear and maintain audience interest (5 points)

**Step 2**

This section lists the Project 1 presentation requirements. Each group will prepare a formal, 10-minute presentation that covers the following points.

* An executive summary or overview of the project and project goals:
  + Explain how the project relates to the industry you selected.
* An overview of the data collection, cleanup, and exploration processes:
  + Describe the source of your data and why you chose it for your project.
  + Describe the collection, exploration, and cleanup process.
* The approach that your group took to achieve the project goals:
  + Include any relevant code or demonstrations of the application or analysis.
  + Discuss any unanticipated insights or problems that arose and how you resolved them.
* The results/conclusions of the application or analysis:
  + Include relevant images or examples to support your work.
  + If the project goal was not achieved, discuss the issues and how you attempted to resolve them.
* Next steps:
  + Briefly discuss potential next steps for the project

**# Convert to floats and store Poverty Rate and Bank Count as x and y valuesx\_values = census\_data\_complete["Poverty Rate"].astype("float")y\_values = census\_data\_complete["Bank Count"].astype("float")# Run linear regression(slope, intercept, rvalue, pvalue, stderr) = linregress(x\_values, y\_values)regress\_values = x\_values \* slope + interceptline\_eq = "y = " + str(round(slope,2)) + "x + " + str(round(intercept,2))# Plot scatter plotplt.scatter(x\_values,y\_values)# Plot regression lineplt.plot(x\_values,regress\_values,"r-")plt.annotate(line\_eq,(6,10),fontsize=15,color="red")# Label plotplt.xlabel('Poverty Rate')plt.ylabel('Bank Count')# Print r square valueprint(f"R squared: {rvalue\*\*2}")# Show plotplt.show()**