

Course Project Introduction

Your course project is an opportunity to showcase your analysis skillset in a collaborative setting. The final result will become part of your analysis portfolio that you may share online with potential employers. The project goals are:

1. Practice data management, visualization, and analysis techniques from the course.
2. Report your [analytical story](#) in a markdown file posted to a GitHub repository.
3. Record a five minute presentation of your findings to a business decision maker.

Required Tasks of the Analytics Story

Your successful project will accomplish the following major tasks:

1. Propose a research topic (Option 1 or Option 2)
2. Create a GitHub repository and establish best practices for team collaboration.
3. Demonstrate merging of multiple data sources
4. Visualize data using Tableau, R, Python, or a combination
5. Generate meaningful summary statistic (KPIs) of the data
6. Submit draft of progress at Checkpoint 1 and Checkpoint 2.
7. Summarize your findings in a short video presentation.
8. Publish a detailed, well formatted markdown report of your analytical story to your GitHub repository. Report requirements are outlined below.

Research Topic

Option 1: Propose a research question for approval by the instructor. Propose a well scoped analytical approach that exploits the tools learned in class. The proposed project may be done entirely in one tool, or in a combination of tools. Specify multiple data sources and how they can be merged together to complete the analysis. The project should have real world application to a decision maker.

Option 2: Select a research question related to the course data. You will be combining the [data on sales of liquors in Iowa](#) with Iowa's demographic and economic data available through [American Community Survey](#). Such merged data will allow you to search for any notable patterns in how liquor sales in Iowa's counties, cities and zipcode areas relate to socioeconomic

factors such as population density, income, education and employment. Details for this project option are in the Annex of this document.

Report Content

As graduate students, you are given a wide level of autonomy in your analysis and report writing. There are however, several requirements that allow for fair grading while simulating a real world analysis project. Your project should tell an analytical story, however, your narrative should remain unbiased to any particular decisions.

1. **Introduction (5pts):** Provide context regarding your decision maker, organization, and overall decision climate. State your research question. Explain how policy decisions will affect your organization and the broader community.
2. **Data summary (5 pts):** Provide a short description of the nature of the provided data set and explain how these characteristics affect your analysis methodology. Summarize the data set with basic descriptive statistics as applicable.
3. **Data analytics (50 pts):** Provide data analytics that add clarity to the research question. Thoroughly discuss insight obtained from your visualizations and analysis of aggregated, data. Suggest an *excursion*, and provide supporting analysis. Plots should be well formatted according to best practices learned in class. Discuss the advantages and challenges of performing analysis in your chosen software tool.
4. **Conclusion (10 pts):** Summarize the analytical methodology and provide closure to your analytical story. Succinctly answer the research questions. Highlight the limitations of your findings and recommend future work. Do not make policy recommendations here.
5. **Policy recommendation (10 pts):** Introduce a specific policy decision that your decision maker is facing. Provide a data driven recommendation for their decision. Explain probable first and second order effects of the recommendation. Explain the benefits and risks of the recommendation.

Project Checkpoints

We will have two ungraded project checkpoints for feedback throughout the course.

Project Topic Approval: Each group will submit their proposed project on Canvas by assigned date.

Checkpoint 1: Publish a markdown document to your GitHub repository discussing your progress on the eight major tasks. Comment on the challenges and victories of collaborating on GitHub. Report the individual contributions of each team member. The latest working files should be pushed to the GitHub Repository. Name the GitHub branch “Team_<#Checkpoint_1.0”

Checkpoint 2: This draft report should contain all the components of the first draft. Strive to have significant progress on each major tasks. Show progress on your research question and an excursion. Name the GitHub branch “Team_*#Checkpoint_2.0*”

DevOps with GitHub

DevOps is a methodology used to collaborate on tech products such as software or analysis projects. The application of beginner level DevOps is one of the stated project goals. Therefore, you will collaborate with your team in a GitHub repository. Use your GitHub repository to post draft projects and the final project. Each team's repository will be public facing to allow other teams to view it.

You should strive to learn basic git functions such as push, pull, add, branch, ect. There are several ways to manage your repository depending on your comfort level. I recommend [GitHub Desktop](#) for beginners and [gitbash](#) for intermediate users. This [git cheat sheet](#) may be enough to get started. Otherwise, you may watch the first several videos of the DataCamp course posted in course resources.

Version Control (10 pts). Maintain your work on the GitHub repository. Commit (save to git) your work frequently. Create named branches for draft checkpoints and final draft. Comment on the challenges and victories of collaborating on GitHub. Novice users should save backup copies of their work on their local computer.

Repository Functionality (10 pts). The repository should be well organized with clear titles of directories and files. Your repository is the public facing home of your work. Include your report as a markdown file. You are welcome to add formatting to your markdown file as long as it enhances the readability on any web browser. The instructor should be able to clone (or download) the repository to their computer and rerun your analysis. Therefore, all required files and data should be self-contained within the repository.

There will be a short learning curve for new git users, but it will be well worth it for your career in econ and analytics.

Annex:

Project option 2 - using provided course data

Groups electing a pre-scoped project should use the following data sets and select one of the following research problems.

Data

All data files for the project are located inside a single archive "project data zip" available through Canvas. The table below summarizes the content inside each file:

File	Description
------	-------------

1	project.sales.zipcodes	Average annual liquor sales per zipcode
2	project.sales.cities	Average annual liquor sales per city
3	project.sales.counties	Average annual liquor sales per county
4	project.acs.counties	ACS data on Iowa counties
5	project.acs.cities	ACS data on Iowa cities
6	project.acs.zipcodes	ACS data on Iowa zipcodes

The first three files contain the same data on liquor sales in Iowa over 2012-2016 that you have already worked with in other assignments, except this time it contains full range of liquor categories. The variables in those files are as follows:

File		Variable	Description
1	project sales zipcodes	zipcode	Zipcode
		category	Liquor category
		sale dollars	Average annual cost of liquor sold in dollars
		sale volume	Average annual volume of liquor sold in liters
2	project sales cities	city	City name
		category	Liquor category
		sale dollars	Average annual cost of liquor sold in dollars
		sale volume	Average annual volume of liquor sold in liters
3	project sales counties	county	County name
		category	Liquor category
		sale dollars	Average annual cost of liquor sold in dollars
		sale volume	Average annual volume of liquor sold in liters

Files 4-6 contain ACS data on various economic and demographic variables across Iowa geographies. Each file contains a unique variable that defines the geography (zipcode, city, or county). The following variables are common across all files:

Variable	Description
high.school	Percent of population, high school graduate or higher
bachelor	Percent of population, bachelor's degree or higher
unemployment	Unemployment rate, population 16 years and over
income	Median earnings, dollars

population	Total population
pop.white	Total population, white
pop.black	Total population, black
pop.indian	Total population, American Indian and Alaska Native
pop.asian	Total population, Asian
pop.hawai	Total population, Native Hawaiian and Other Pacific Islander
pop.other	Total population, other single race
pop.multi	Total population, two or more races

Pro Tips: If using Tableau, use 3 workbooks. When merging ACS and annual sales data, make sure to use full outer merge, as it is possible that not all ACS geographies have recorded liquor sales and/or not all sales geographies have corresponding matches among ACS geographies

Research Questions

If not pursuing a self selected topic, teams may choose from one of the following:

- 1:** What is the distribution of per capita sales across geographies? What are the ranks of top 10 geographies for per capita consumption across every liquor category? Are there any outliers that have high per capita sales in only one specific liquor category?
- 2:** Are there pairwise patterns in total/per capita sales across geographies? Are there correlations between average sales per zipcode and average sales in corresponding city? Does the pattern differ across liquor categories?
- 3:** Do geographies with higher median income consume more alcohol in total? What about per capita? Does it depend on liquor category?
- 4:** Does employment affect what liquor categories are sold most? Is the pattern the same as the one for median income?
- 5:** Are there any preference patterns for liquor consumption among different races? What are most popular liquor categories among geographies with highest share of minorities?
- 6:** Is there any notable difference in liquor sales across communities with varying levels of education?