**Project Name:** Retain Alabama

**Project Team:** Michael M Gathara, Tanner L Tavel, Michael D Hurt

**Senior Data Fellow:** Daisy Homolka

*For senior data fellow: Please finalize the below project plan for your project team and provide project-specific direction/notes in the subsequent pages. Upon completion, please submit to Dr. Lauren Rast (*[*laurenr@uab.edu*](mailto:laurenr@uab.edu)*) by Tuesday, June 15.*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Action Item | Week | | | | | | | | | | |
| 6/7 | 6/14 | 6/21 | 6/28 | 7/5 | 7/12 | 7/19 | 7/26 | 8/2 | 8/9 | 8/16 |
| Define the project goal and execution plan  *Complete project plan, complete prep work, form teams and communication schedule* |  |  |  |  |  |  |  |  |  |  |  |
| Understand the problem and project goal  *Review project scope of work, reading materials and datasets* |  |  |  |  |  |  |  |  |  |  |  |
| Define specific data and tool(s) needs  *What data do you need? How and where will you retrieve it?* |  |  |  |  |  |  |  |  |  |  |  |
| Get and understand the data  *Gather the data from multiple data sources and define the various data points* |  |  |  |  |  |  |  |  |  |  |  |
| Explore and clean the data  *Clean and format the data, preparing it for analysis* |  |  |  |  |  |  |  |  |  |  |  |
| Find insights  *Conduct exploratory analysis as you work to answer questions in the project scope* |  |  |  |  |  |  |  |  |  |  |  |
| Validate data  *Did you answer the right question? Did you interpret the data correctly?* |  |  |  |  |  |  |  |  |  |  |  |
| Visualize the data  *Tell the story through your project deliverables (i.e., interactive dashboard, report, presentation, etc.)* |  |  |  |  |  |  |  |  |  |  |  |
| Package project deliverables  *Finalize report and presentations and prepare to present* |  |  |  |  |  |  |  |  |  |  |  |
| Present findings  *Present projects to clients at Demo Day* |  |  |  |  |  |  |  |  |  |  |  |

**Reference** [Understanding the Lifecycle of a Data Analysis Project (northeastern.edu)](https://www.northeastern.edu/graduate/blog/data-analysis-project-lifecycle/)

**Reference** [7 Fundamental Steps to Complete a Data Analytics Project (dataiku.com)](https://blog.dataiku.com/2019/07/04/fundamental-steps-data-project-success#:~:text=%207%20Fundamental%20Steps%20to%20Complete%20a%20Data,next%20data%20science%20step%20is%20the...%20More%20)

**Project Plan Notes**

***For senior data fellows:*** *Please provide high-level direction and project-specific notes, including attachment references, within each section corresponding with the project plan on pg. 1.*

1. **Define the project goal and execution plan**

The final deliverable will be a summary report and a PPT that provide analysis and visualization of the survey data to uncover the most important factors for retaining graduates in Alabama.

Students will test and explore several predictive models, likely focusing on multiple regression

*OPTIONAL*

*Use the qualitative discussion board data to conduct a sentiment analysis. Compare qualitative and quantitative results.*

Weekly meeting will be on Tuesdays at 2:30pm.

1. **Understand the problem and project goal**

Students should look over the UAB Survey Data, Statewide Data, Qualitative Data, and Survey Information to get a general idea.

Read Jupyter Notebook that has an overview of packages and potential machine learning models (Daisy will provide)

1. **Define specific data and tool(s) needs**

*DATA:*

UAB Survey Data

Statewide Data

Qualitative Data

Survey Information

*TOOLS:*

Python

Jupyter Notebooks

Various ML/data science packages to be defined later

1. **Get and understand the data**

Students should understand where the data comes from, how it was collected, and what do all the variables mean.

Understand margin of error, potential sampling bias etc.

1. **Explore and clean the data**

Exploratory Data Analysis (variable types, distributions, values)

Explore ideas for feature engineering and/or factor analysis

Investigate relationships between the variables

Transform the data so it can be used for regression analysis including creating dummy variables

*OPTIONAL*

*Transform the qualitative data into something usable for potential sentiment analysis*

1. **Find insights**

Train, cross-validate, and test predictive models, focusing on multiple regression techniques

*OPTIONAL*

*Sentiment analysis on qualitative data*

1. **Validate data**

Select the best model and clearly interpret and justify the results

*OPTIONAL*

*Clearly interpret and justify the results of the sentiment analysis on qualitative data*

1. **Visualize the data**

Select and fine tune the best visualizations for both descriptive statistics and the final predictive model

Write a report clearly explaining the insights from the data and the results of the predictive model. The report should be written for a business audience, not a technical one. It should be easy to read, clear, and well-designed.

1. **Package project deliverables**

Finish formatting and writing the report

Create a PPT that can be used to present the report, both by students at demo day and later by EDPA/UAB to share the findings with stakeholders.

1. **Present findings**

Present PPT at Demo Day