

**Group Project**  
**Data Programming in Python**  
**Business Analytics Graduate Program**  
**BAIS:6040**

The goal of the project is to learn the python language by applying data science techniques in the context of creating an operational predictive model. The course-long group project will incorporate gathering, cleaning, transforming, and exploring data; followed by fitting, tuning, and building a model that is operationalized in a python Flask application. The emphasis is on gathering, cleaning, transforming, and exploring the data, **not** on the accuracy of the predictive model.

The project includes a checkpoint where the group will submit a Jupyter notebook with the group's progress to date. The expectation in the final presentation is a working Flask application that at a minimum returns JSON.

Please follow these steps to achieve this goal:

### **Problem Identification**

Identify a business topic that interests your group and which you can easily collect a reasonable amount of data to make a classification or regression prediction.

It could be a situation at work or something involving a hobby. However, this is a business analytics course, so finding a business-related topic is preferable. If you need some ideas, let the instructor know. If the data comes from work or another private source, please make sure that you can use the data. Ideally, you should also be allowed to let the instructor see the raw data and to summarize the data for your classmates, but compromises can be worked out with the instructor if necessary.

Do not choose one of the topics that have already been covered in the projects of other courses such as Data programming in R. Just redoing your previous R project in Python is not an acceptable idea.

### **Goal Determination**

Determine a question that you would like to investigate about the situation using the data you plan to collect.

The question should be interesting in the sense that: (i) it does not have an obvious answer; (ii) the data is necessary for answering the question and creating a predictive model; and (iii) the

answer will provide insight into the situation. For example, suppose you were interested in analyzing Des Moines real estate. You might ask, “What factors are predictive in determining Des Moines residential real estate values?” Then use those factors to build a model that predicts specific property values.

## Core Analysis and Modeling Building

Follow these generic steps to complete the core work of the project using Python and Jupyter Notebook:

- Gather the data and import it into a Jupyter notebook.
- Transform, clean, and prepare the data for analysis.
- Analyze the data to answer your question fully. Your analysis should include viewing the data graphically.
- Build a predictive model (either classification or regression) using the data. You should calculate and report the accuracy of the model, but the emphasis of the project is NOT on the accuracy of the model.

**During these steps, you must also write and use at least one Python user-defined function to somehow help your work.**

For example, if you would like the option of querying your database in different ways, you could write a function that accepts different options and outputs a customized SQL query. Or if your data cleaning includes handling text data in a particularly tricky way, you could separate that part into a function leaving the main Python code easier to read. Or if you would like to have the option of preparing different graphics based on different options, you could write a function to enable this. Overall, the function can be used to assist your handling and analysis of the data, or it could just an “add on” function that is auxiliary to your handling and analysis of the data.

Throughout the process, use a single Jupyter notebook for data gathering, cleaning, transforming, and exploring and **another** notebook for modeling. The notebooks should contain all the commands necessary to reproduce your analysis, including importing the data and producing any graphics. Please comment all your code. Also, include the pip install commands in the notebooks if you are using any external Python packages, so the instructor can easily build the same environment.

## Flask Application

The Flask application will need to be created in a .py file. The instructor will provide a template for creating this file. Your .py file should be submitted along with any presentation materials you create for the demonstration.

Have one group member submit by the deadline a **single compressed Zip file** containing both Jupyter notebook files (.ipynb), the Flask application (.py file), and any data sets used. **Note that**

**you do not have to make a separate document as a report. Everything should be contained in the Jupyter notebook(s) and the .py file.**

You will also need to provide at least one observation the instructor can use to test your application.