TEAM MEMBERS:

Brinda Sapra - brinda
Julián Varón Parra - jvaro
Diego Alejandro Martinez Cruz - dmartinez
Pedro Antonio Ramonetti Vega - pramonetti

Project Description

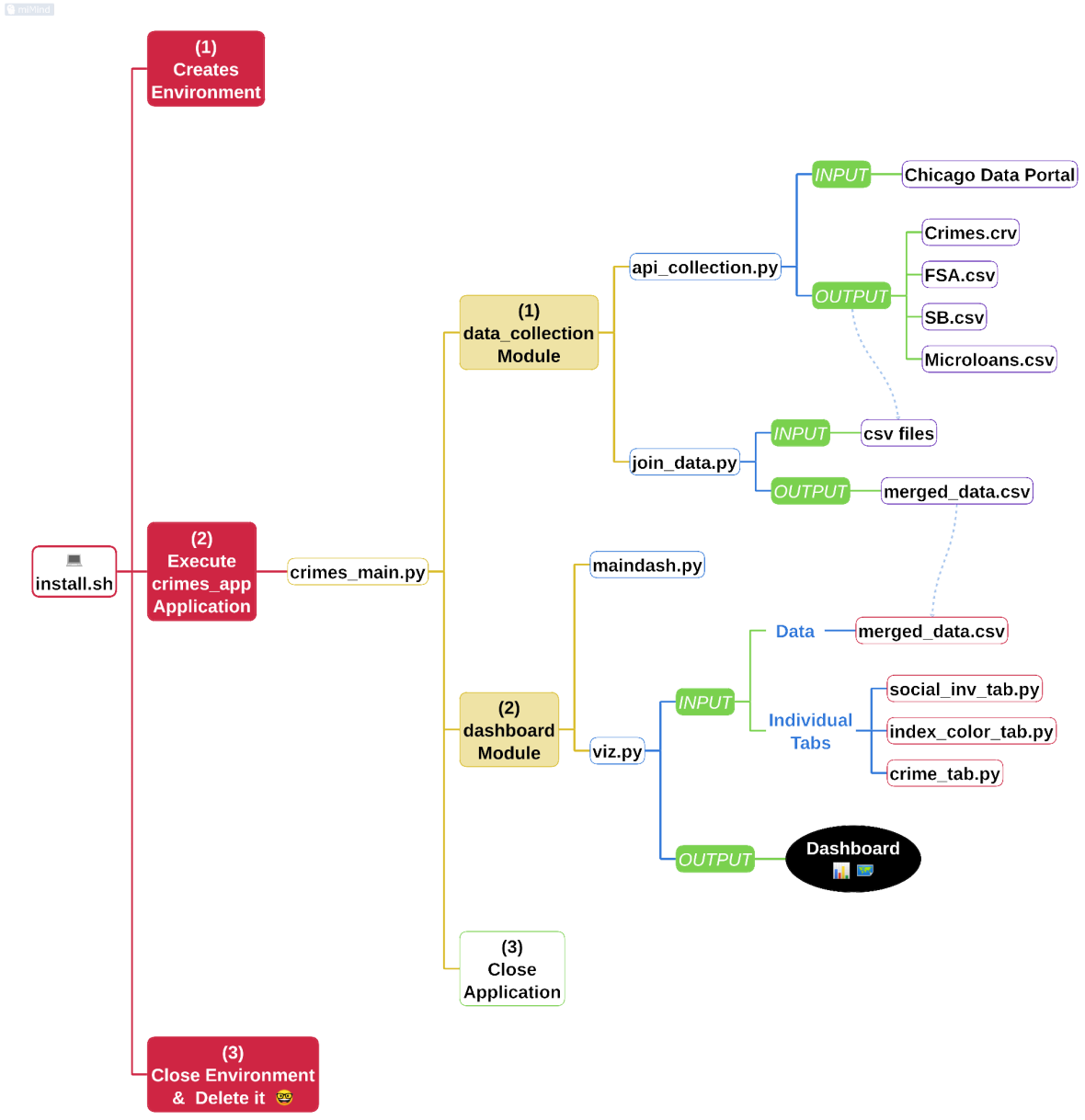
The purpose of this project is to explore a relationship between the incidence of crimes in all the wards of Chicago and the amount of social investment that is directed towards a ward. The hypothesis is that providing more investment through social programs, such as supporting microbusiness and creating social recreation spaces, is negatively related with the number of crimes reported within the ward i.e. wards with the higher social investment report a lower number of crimes.

We model a linear regression at ward-level:

1. Dependent variable - total number of crimes
2. Independent variables - 3 proxy indicators for social investment:
    - Number of Family Support Agencies
    - Total amount of investment through the Small Business Improvement Fund
    - Number of loans given by the Chicago Microlending Institute

The residuals (error term) from the regression is used to build an index between -1 and 1. A negative value of error indicates that the ward's number of crimes are less than the regression's predicted number of crimes, so there is "overinvestment" of social capital in that ward. A positive value of error indicates that the ward's number of crimes are more than the regression's predicted number of crimes, so there is "underinvestment" of social capital in that ward.

Structure of the software



Tasks Distribution

Our team collectively considers that all of us contributed equally to the project.

Task	Description	Responsible
Data, gathering, merging and cleaning	In charge of downloading the several datasets, and merging them using “wards” as keys.	Pedro Ramonetti & Julian Varon
Regression Analysis	Create a linear regression module that takes the crime outcomes, as a dependent variable, and the explanatory variables related to social investment. From this model, the security dependent variable for each ward would be predicted, and the estimated error will help us to see the position of the security-investment relationship of each ward in comparison with the average ratio. The errors will be displayed on heatmap (high error will be shaded darker than low error for each ward).	Diego Martinez
Visualizations	Exploratory analysis by doing cross-tabs of different variables to see which variables can be made available to the user to select from a dropdown (across all datasets) to explore meaningful relations to add context to the regression results	Brinda Sapra & Julian Varon
GIS Implementation of heat map	Use shapefiles to draw the map boundaries for Chicago and use	Brinda Sapra & Diego Martinez & Pedro Ramonetti
Application	Connecting all previous work to the user command-line interface	Pedro Ramonetti

Interaction

The project has its root in the install.sh file. Once this file is run, it will fire up a virtual environment and install the required packages using requirements.txt to run the application. After this, the application will display a menu of 3 options for user to select:

1) Download data from API: This choice will call the download file and merging files, replacing the current files in the data folder.
2) Run the visualization: This choice will open a dashboard, in which the user can interact with the number of crimes at ward level, the variables for social investment, and the heatmap of the index.
3) Close the app: This choice will close the application, close and delete the virtual environment.

Project Goal and Results

Our objectives with the development of the projects can be classified in three main categories: first, from a general perspective, our goal was to develop a tool to understand the relationship between social investment and crimes in Chicago. Second, we wanted to explore visualization tools such as Dash, Plotly, GeoPandas, and delve into techniques for gathering, merging, and cleaning data using Sodapy (API for Chicago Open Data portal) and Pandas. Finally, through the project's development, we aimed to develop a teamwork methodology to efficiently integrate the different pieces of code created by each group member and integrate the various modules into the application using a package structure.

We accomplished the above objectives and fully covered what we proposed in the previous deliverables. As we explained in the project description, we could map the specific relationship between social investment and crimes. We also achieved a better understanding of Dash and Pandas functionalities, API's and virtual environments, and the different modules' integration in our application.