

AegeanSEE: Visualizing Investment Opportunities in the Aegean Islands
Amir Shahatit, Khalil Joseph, Parsa Attari, Rami Shahatit, Yash Kirloskar

Intro:

Our project was to create a user interface heat map of the Greek Aegean islands, focusing on the features that would make a given island a good investment - for tourism, infrastructure, and so on. Our platform is designed to be useful for the researchers that we coordinated with at the University of the Aegean, aiding their efforts in completing economic analysis. We believe that our use of Scikit-learn, Pandas, and feature selection to develop a user-friendly tool will benefit the efforts of the researchers abroad.

Problem Statement, Journey, Tools, What Did/Didn't Work:

The Greek Aegean Islands offer a great opportunity for economic growth for the country, but the information and modeling has not been completed yet, so the land lays tucked away, overlooked for tourist, real estate, and other ventures for many years. We started by trying our best to familiarize our team with the landscape and environment of the problem at hand, and sought out to create an FDI prediction model using data from other countries of similar GDP. We then, after speaking with the team in Greece, had to pivot our current work towards something that could be utilized as a time-sensitive heat map, even though we didn't have sufficient data a lot of the time. Through our consideration of many features, we were able to transfer this data into a user interface heat map that would display the relative ripeness of investment for a given island. One of our largest challenges was mediating what to build, what we *could* build, and what the research team wanted, as all of us came from different backgrounds and had different approaches to this project. We worked past these conflicting points of view by trying our best to put the end-user at the forefront of our project, the team in Greece. We paired this with weighing how much value we could provide them give the time constraints, and came to the conclusion of the ML-based heat map. We leveraged tools such as Sklearn, Pandas, and geopanda/foulm to be able to put forth a product that had the right direction, was simple to understand, and met the needs of the offshore team.

Full-System Solution/parts it contains/why valuable/future improvements:

The data provided by the University of the Aegean first required clean up in order to make it uniform across all regional units or islands. Due to time constraints and the fact that this process had to be done by hand, we completed this for 5 sets of the provided data. A future improvement given our knowledge of this portion of the project would be to automate it as much as possible and include the option to build to selection given three general regions (which are larger than a single island). This would be helpful because a great amount of the data came in like this, therefore making it hard to use with the island comparison model we created.

After this point, we used Pandas to read in the csv file, then normalized all the data in order to make the variance smaller. This was done in order to make finding a variance threshold easier. In

the future, we could have utilized automation of the variance threshold and letting users choose for themselves how many features they wanted to keep.

Finally, we used Scikit-learn's variance threshold - a feature selection module - to remove features whose variance was less than 0.117, which left us only 10 features that would be input into the heat map user interface. This was ultimately of great importance to developing a good heat map, because our visualization would only be as good as the features we selected, and we utilized variance as a key variable because it proves to be a differentiator.

At this point, we had to create the heat map so that we had a proper representation of our data. We used GeoPandas and Folium to take in the inputs from the csv file, arrange, and plot the data within our notebook. Next, we matched the names of the islands to longitudes/latitudes of the segmented data. Then, we formatted the data to be able to pass it into the heat map plug-ins. This all allowed us to create "hot" and "cold" zones for investment and represented on the map.

Team Contribution:

Our team as a whole worked well together. We coordinated with each other using mediums like text messaging and Facebook Messenger. We scheduled meeting times as a whole and with individual team members in order to be most productive. Upon our first couple of meetings we needed to learn more about each other's skills and strengths in order to properly distribute tasks. Once we figured that out, we began assigning tasks. We had to adjust to the late addition of a member and an even later dropping of a partner over the course of the semester. Ultimately, it helped keep the team balanced and agile in the long run, as the team grew bigger and smaller which kept us on our toes. Together, we brainstormed ideas and bounced some implementation details off each other. Individually, our team worked cross-functionally, helping one another, but we segmented the portions of the project to keep deadlines and requirements organized.

Yash worked on creating the machine learning algorithm implementation, which involved everything from cleaning the data to feature selection. Parsa focused on the UI that would link the csv file returned from Yash's algorithm to a map of the islands and represent the data in a heat map format. Amir worked on the presentation for the low tech demo, presentation slides in general, and helped with the written portions. Khalil coordinated with our mentors/professors/TAs in order to ensure organization, the written portion, and researched/architected ideas that we could input into our machine learning algorithm. Rami helped with presentation structure, and also researched/architected ideas that went into our machine learning algorithm.

Mentor Experience: Our team worked with Professor Yannis Charalabidis and Zoi Lachana primarily. We greatly appreciated their guidance and deep knowledge of the subject matter, taking into account their research and opinions when we were deciding on the direction of our

project. That being said, it was difficult coordinating with the team at times, and aside from a handful of Skype sessions and emails, it was difficult to keep in contact with them while figuring out the direction of our product. We initially tried to pool in data from countries similar in GDP to Greece to possibly use classification to characterize foreign direct investment opportunity, but the team in Greece had a different plan in mind, which was more of a ‘temporal’ view of the heat map, that recommended when to invest in a certain island. The issue with this method was that there was not a wide variety of resources regarding financial reports of individual islands, along with tourism and population data.