1. **Title of the Idea:** “Machine Learning for Small Business Merchandising”

2. **Mentor names:** Joe Garner and Ali Zaidi

3. **Each author's name, affiliation, city, state, and country of residence, and email address:**

Sephora Jean-Mary, Student, Miami, Florida, United States, [sjean149@fiu.edu](mailto:sjean149@fiu.edu)

Charlie Ramirez, Student, Miami, Florida, United States, [crami097@fiu.edu](mailto:crami097@fiu.edu)

4. **Scientific/engineering interests or areas of expertise of the author(s):**

Sephora Jean-Mary:

*Areas of expertise*: Front-end developing using tools such as HTML/CSS/Javascript.

*Interests*: Building websites

Charlie Ramirez:

*Areas of expertise:* Front-end developing using tools such as HTML/CSS/Javascript, cross-platform app development using Flutter.

*Interests:* Learning as much as possible about different technologies.

5. **Project description:**

Online commerce has generated oceans of data on consumer behavior and trends in fashion and consumer needs. Unlike their larger counterparts, small businesses lack the time and resources to obtain and analyze this data, generate domain-specific forecasting models, or monitor their brands during new product launches. What they need is quick access to this data for little effort and cost.

In this project, we will demonstrate the ease with which any small business merchandiser can access online product information and generate meaningful visualizations.

We plan to build a web scraper and scrape product information from Amazon product pages and other related online social media entities such as Twitter, Reddit, and various blogs. In short order we will extract scraped data such as descriptions, ratings, reviews, prices, and other features, extract the data to a simple database, analyze the data, and produce meaningful visualizations easily accessible on an interactive website (e.g., rankings & features correlations, competitive trends, price forecasting, consumer behavior profiling). For the scraping, database, and presentation services, we will use Opensource technologies (Python, Jupyter notebooks, SQL, JavaScript); for the Machine Learning algorithms, we will use popular packages available at sites such as Python.org, scikit-learn, and GitHub.

We hope that others can build upon this work and make data science more accessible to the moms and pops.

6. **Why this idea and what are the benefits we can expect from this product?**

*We know that machine learning and AI will form a great part of our society’s future. If businesses are to survive and thrive in an AI dominated market, they must be able to adapt to the changes that are occurring. This is the goal we aim to achieve with this project. Large corporations such as Amazon are dominating the retail market. Articles like* [*this*](https://techcrunch.com/2018/07/13/amazons-share-of-the-us-e-commerce-market-is-now-49-or-5-of-all-retail-spend/) *one allow us to see the extent to which this is true; almost half of all e-commerce sales happen through Amazon (5 percent of the entire retail market). This creates a great opportunity, which we aim to explore, for businesses to increase their sales.*

*Additionally,* [*studies*](https://www.thedrum.com/news/2017/03/27/online-reviews-impact-purchasing-decisions-over-93-consumers-report-suggests) *suggest that online reviews have a great impact into the decision of whether to buy a product. Customers* [*expect*](https://www.statista.com/statistics/1019495/online-shoppers-expectations-product-reviews-in-the-us/) *products to have many reviews, as these give them an idea of the quality of the product. The benefits of taking reviews into account, such as* [*lower return rates*](https://footwearnews.com/2019/business/retail/product-ratings-online-retail-sales-1202760878/)*, is something we will explore in this project. Through technologies such as web scraping and machine learning, we aim to bring a product that will allow businesses to survive and thrive in the AI dominated years to come.*

7. **Proposed schedule of activity**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sprint** | **Date(s)** | **Tasks** | **Resources** |
| Sprint 2 | Jan 27 – February 9 | -Understand project requirements  -Design architecture of back end and front end  -Start coding the Amazon web scraper (use Python) | Python Scraping/Crawling   1. <https://www.scrapehero.com/tutorial-how-to-scrape-amazon-product-details-using-python-and-selectorlib/> 2. <https://www.promptcloud.com/blog/tutorial-how-to-scrape-amazon-product-details-prices-using-python/> 3. <https://blog.datahut.co/tutorial-how-to-scrape-amazon-data-using-python-scrapy/> |
| Sprint 3 | February 10 – February 23 | -Finish coding web scraper  -Test web scraper, make sure all data elements are being captured and that the dataset is clean  -Begin coding website (in language of choice) |  |
| Sprint 4 | March 2 – March 15 | -Finish testing the web scraper  -Continue coding the website  -Start creating visualizations for the website (in language of choice) | Python Visualizations (if that’s what we want to use)   1. <https://mode.com/blog/python-data-visualization-libraries/> 2. <https://mode.com/blog/python-interactive-plot-libraries> |
| Sprint 5 | March 16 – March 29 | -Continue coding the website  -Finish creating visualizations  -Begin using ML models to predict various characteristics of data (predict rating, predict # comments, etc. based on discussion) | Python Modeling of Amazon Data   1. <https://towardsdatascience.com/predicting-sentiment-of-amazon-product-reviews-6370f466fa73> 2. <https://medium.com/jbencina/part-1-predicting-amazon-review-ratings-with-text-analytics-in-python-fa7c14e91464> 3. <https://pdfs.semanticscholar.org/959b/6d911898ac04dcc706d3d326142d9bbf454b.pdf> 4. <https://github.com/drewmassey/amazon_reviews> |
| Sprint 6 | March 30 – April 12 | -Continue coding website  -Finish ML models  -Start visualizations of model outputs (predicted vs actual outputs) |  |
| Sprint 7 | April 12 – April 26 | Test all components, including scraper, website, models, and data. Continue adding design details and making updates where needed. |  |
| Final Deliverable Due | April 27 | Turn in deliverable |  |

8. **Up to three key words describing the Big Idea;**

*Merchandising, Sales, Commerce*