That's So Last Year

Group Members:
Ben Joseph Rivera(PM): benrivera@arizona.edu
Braydon Polly-Holguin: bpollyholguin@arizona.edu
Louise Ann Lalescu: lalescu@arizona.edu
Sophia Mikulewicz: smikulewicz@arizona.edu
Ryan Lenhart: lenhartr@arizona.edu

Description:

It seems that everything around us is going up in price. Gas, groceries, bills, and clothes are all skyrocketing in price. With so many applications already comparing prices for many of these, clothes seem to be an untouched niche. Some sites can compare prices of a similar item, but the extent of its capabilities. Applications like Google Shopping already do this. However, our proposed website, **That's So Last Year**, will advise shoppers on what stores would be the best to shop for given certain circumstances. Our website will visualize five fashion websites based upon certain criteria: price of specific types of clothing, average costs of specific types of clothing, minimum and maximum prices of clothing, and the total number of choices of each specific clothing.

Fashion is an industry that creates a lot of revenue. In order to do so, companies or fashion industries need to understand what their consumers are willing to wear. On top of advising what fashion websites for people to visit for certain aggregations, we want to include an analysis on how consumers' tastes correlates with what is being sold on fashion websites. Therefore, if one website has a surplus of shorts compared to others, then perhaps starting a search for the perfect pair of shorts should start there. These websites will range from fast, sustainable, and department stores.

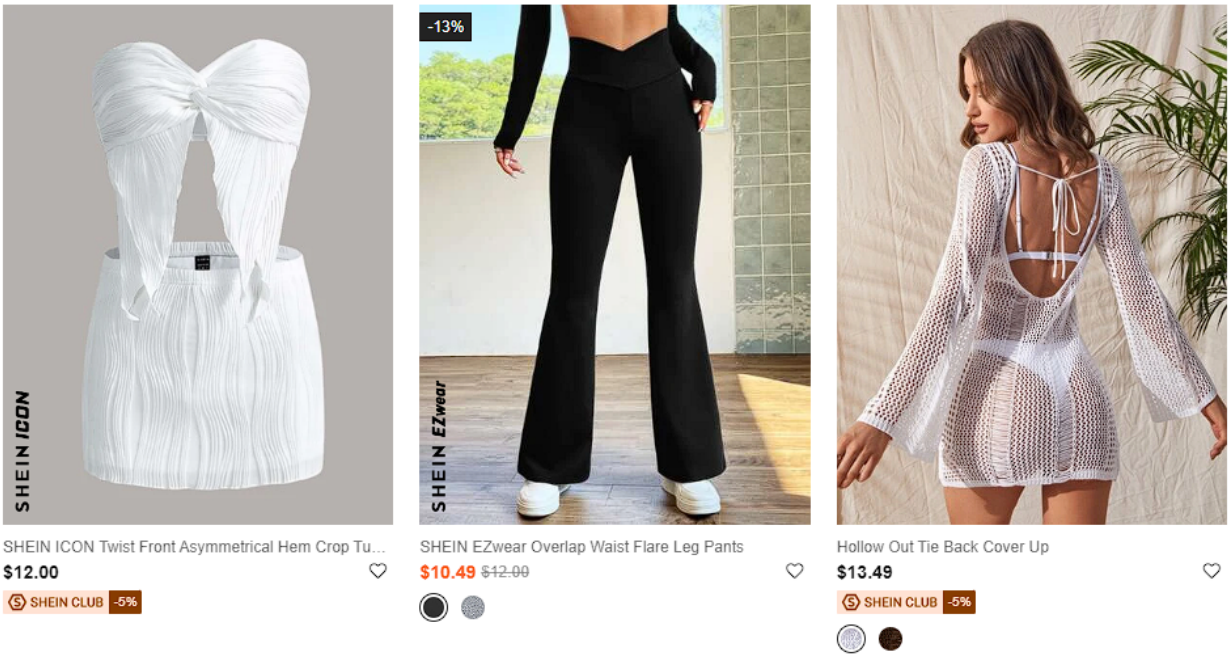
Skills:

Ben - Website building, MSAccess, Tableau, Excel
Braydon - Python, Excel, Tableau, Data Scraping
Louise - Python, Web Scraping, Data storage, Data Engineering, Tableau, Data Analysis
Sophia - Python, R, SQL, Web scraping, Data analysis, Data visualization,
Ryan - Python, HTML, CSS, Data Engineering, Some Data Science

---

Data:

Most of the data that will need to be scraped are on websites that do not have static data. The data is dynamic and requires a much more tedious approach to scraping. We have opted to use selenium for such cases. However, this can be complicated by security systems such as Cloudflare. When trying to use selenium to crawl pages, cloudflare would block any requests from scripts. Other websites like FashionNova allow us to scrape via BeautifulSoup. The primary data that we plan to scrape from websites will be the price of each clothing and the description of said clothing. An example of dynamically scraped data would be like shein:



The websites, as of this submission, that will be scraped are as follows:

Fast Fashion:

-Shein: (Name of product, price)

-FashionNova: (Name of product, price)

Department Stores:

- Khols: (Name of product, price)
- Macy's (Name of product, price)

Sustainable:

- ThereFormation(Name of product, price)
- Aday (Name of product, price)

The data that we plan to scrape from pinterest will depend on data gathered from the websites listed above. The pinterest API allows us to search for certain keywords. The keywords with the

greatest yearly changed are the ideal candidates for our correlation analysis: For example, if we keyword search on Pinterest for dresses, we got the following results:

Keywords	Weekly change	Monthly change	Yearly change
prom dresses 2023	-30%	300%	10,000%+
summer dresses	3%	100%	0%
pink prom dresses	-6%	100%	60%
prom dresses long	-30%	200%	80%
prom dress inspiration	20%	200%	2,000%
pink prom dress	-10%	100%	70%
light blue prom dress	-5%	200%	20%
summer mother of the bride dresses	-1%	100%	100%
spring dresses	20%	200%	10%

Therefore, the most likely piece of clothing to select for a correlation analysis would be “Prom Dress(es). If specific keyword searches do not reflect what is being sold on the fashion websites, we will have to elect to correlate by the keyword “Dress(es)” excluding the prom portion.
Pinterest

- Pinterest Trends API scrape of biggest keywords scraped from fashion websites

---

Mentor:

Rich Thompson: Rich is knowledgeable in Python programming. He will be able to assist us in our web scraping process and creation of data visualizations.

Data Ethics:

Question	Generally	Data Breach
Could a user sell drugs or other illegal items on your platform?	N	N

Could a user of your platform engage in sex trafficking?	N	N
Could a user sell class notes or cheat on their homework on your platform?	N	N
Could a stalker use your project to find someone?	N	N
Could your app be used to spy on or track individuals?	N	N
Could your app/software access the camera or microphone and record things without users being aware?	N	N
If someone uses your platform, could they be re-traumatized or have their mental health impacted in some way?	N	N
Could your algorithm promote material that would traumatize or upset individuals?	N	N
Would your users be upset if the data you collect was given to someone else?	N	N
Could a data leak potentially lead to identity theft?	N	N
If your site was hacked, would users of that product potentially lose their job, spouse, or family?	N	N
Should there be an age limitation on your product?	N	N
Could someone use your product to find, contact, and potentially commit elder abuse?	N	N

If the data on your platform was breached, could it be used to blackmail the users?	N	N
Does the existence of your project imply that a particular racial group, gender, religion or other protected category is inherently bad, gross, or unwatched?	N	N
Could your product be used to commit hate crimes against a specific group?	N	N
Does the primary content of your game or algorithm focus on something considered deeply unethical?	N	N
Does your game or software contain race, gender, or other stereotypes?	Y	Y
Could users of your app scam other individuals?	N	N
Is your particular algorithm biased towards predicting correctly only for one race, gender, or other groups?	N	N
Are the users of your project, players of your game, or those being surveyed for your data aware of how their data will be used?	N	N
What are the possible misinterpretations of your results? For example - would a white supremacist or misogynist be stoked about your results if they misinterpreted it?	N	N
Does the use or purchase of	N	N

your data potentially contribute to a dangerous group or regime?		
Could your virtual reality environment cause injury to the user?	N	N
Are your study participants or game players aware that their data will be collected and used?	N	N
Does your game or app contain addictive design elements without benefit to the user?	N	N
Does your survey contain an aspect of compulsion or unusually large incentive, that would command users to take it even if it was to their detriment?	N	N
Could your research outcomes harm an individual or entity?	N	N

18 - The reason we said yes is because fashion tends to categorize items by assigned gender (Male or Female) and does not take into account those that may not identify by assigned genders.

We feel that the purpose of our project does not violate any ethics nor have any ethical concerns. This is because our project is about fashion, there is no immediate effect on a person's life from fashion. The biggest ethical concerns in fashion come at the hands of the shops, and because we have no correlation with the shop we just provide the information, those ethical issues are not in our projects.

Literature/Market Review:

WGSN: https://www.wgsn.com/en - This is a trend forecasting company that accurately predicts trends for the most valuable brands globally in the fashion industry. This company has a TrendCurve+ service that uses their proprietary forecasting algorithms to provide accurate product predictions. They have experts who utilize data sources from social media and catwalk shows to e-commerce, search and consumer sentiment for their forecasts. They also have global expert-generated data, and their AI-powered forecasts are said to produce a 90% accuracy rate. We will be analyzing trends, not predicting them which is done through TrendCurve+.

https://www.fibre2fashion.com/industry-article/83/fashion-forecasting - This article touches on factors that go into trends in fashion and how they constantly change. There are cultural, commercial, psychological, and sociological factors that dictate certain fashion trends. Although our group is stretching to make a predictive model, we are conducting an analysis about consumer research. Fibre2Fashion says that retailers can directly ask what fashion trends are preferred, but turnout for surveys or response to telephone calls can be a challenge for large scale companies. Therefore, with our analysis relying on big data from Pinterest to analyze what consumers are buying, can be a big help in consumer research.

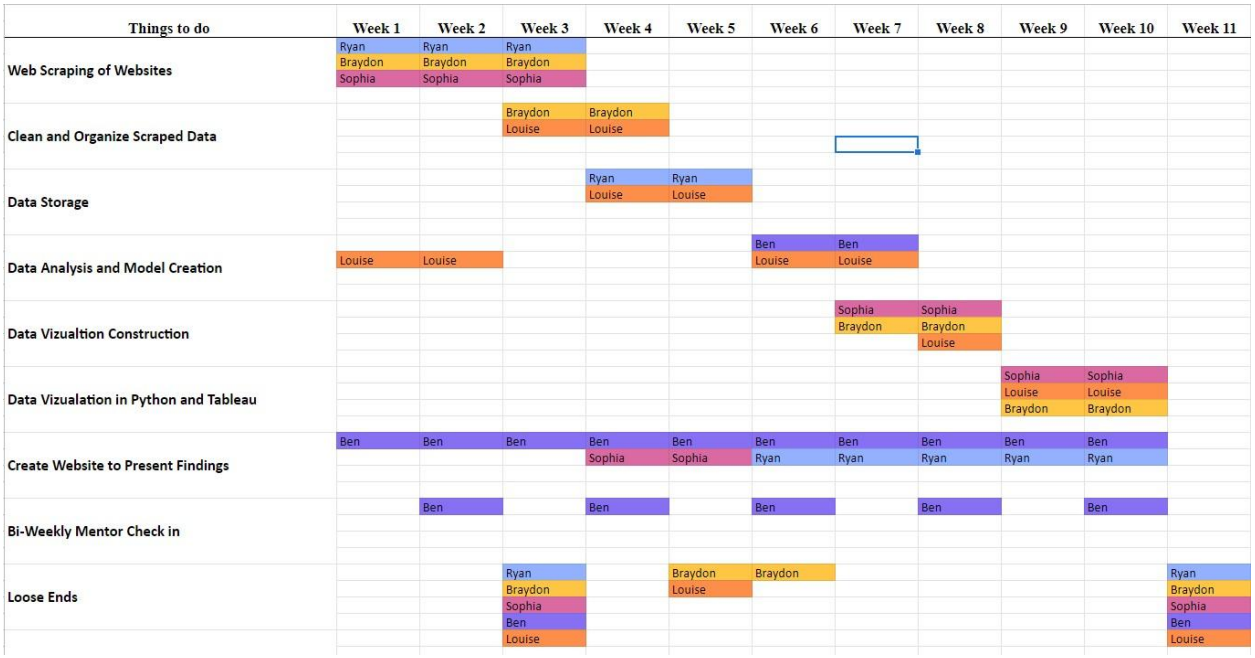
TrendStop: https://www.trendstop.com/ - Not only does TrendStop forecast fashion and lifestyle trends, they also examine why consumers choose certain trends but they do this at a price. TrendStop offers demos and bookings to assist in finding what is fashionable for a user. Though we won't be assisting users, we will be providing users visualization and analysis on popular trends within womens' fast fashion at no cost.

Modacable: https://www.modacable.com/en/ - This source provides predictions up to a year in advance of fashion trends for different markets. These markets include kids wear, mens wear, athleisure, denim, and even jewelry. Though they have these predictions with very specific things such as shape and color they do not provide a source of how they make these predictions. Our project will be different because it will have direct comparison of Pinterest and Shein which can give a strong comparison rather than a baseless prediction.

"Big Data in fashion: trasnsforming the retail sector"- This article from Sirimal Silva focuses on how big data can be utilized in the fashion industry. The article starts off by exploring how big data can be used to predict trends. It then moves on to how big data can be used to minimize waste within the fashion industry. The two primary ways big data can help are by minimizing the number of returns and reducing the amount of waste from overproduction. One of the main issues the article points out is that there is a lack of access to micro-level fashion data when compared to other industries. This lack of publicly available data presents an opportunity for our team to explore the fashion industry through our own data scraping and analysis.

Bloomberg:
https://secondmeasure.com/datapoints/fast-fashion-market-share-us-consumer-spending-data-shein-hm-zara/
Bloomberg goes over the growing market share that Shein has. Shein accounts for 50% of the market share for fast fashion. This source shows that Shein leads the fast fashion industry and as such is a great source for comparing Pinterest trends to the trends seen on Shein. The data we collect from Shein trends will provide the most information on how the trends on Pinterest correlate with actual market data.

Gantt Chart:



Things to do	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11
Web Scraping of Websites	Ryan Braydon Sophia	Ryan Braydon Sophia	Ryan Braydon Sophia								
Clean and Organize Scraped Data			Braydon Louise	Braydon Louise							
Data Storage				Ryan Louise	Ryan Louise						
Data Analysis and Model Creation	Louise	Louise				Ben Louise	Ben Louise				
Data Vizualtion Construction							Sophia Braydon	Sophia Braydon Louise			
Data Vizualation in Python and Tableau									Sophia Louise Braydon	Sophia Louise Braydon	
Create Website to Present Findings	Ben	Ben	Ben	Ben Sophia	Ben Sophia	Ben Ryan	Ben Ryan	Ben Ryan	Ben Ryan	Ben Ryan	
Bi-Weekly Mentor Check in		Ben		Ben		Ben		Ben		Ben	
Loose Ends			Ryan Braydon Sophia Ben Louise		Braydon Louise	Braydon					Ryan Braydon Sophia Ben Louise

---

<u>Compatible Systems:</u>
-Windows and Mac

<u>Minimum Deliverable:</u>
The minimum deliverable will be a locally hosted website with integrated Tableau visualizations and a correlation analysis that will allow users insight to where they may best find clothing for their specific needs.

Expandable Deliverable:

Our stretch goals will involve the following additions to the website and visualizations.

- Color data for listings
- Searchable terms
- Website data filters
- Similar product suggestions
- Interactive Tableau dashboards