Progress Update

**Week 10**

1. **Communicated with Nicole to get feedback**. The following feedback points were noted and refined:

* Reformat all texts on poster to put emphasis
* Retitled findings to sustainable design insights
* Bolded the overview diagram to distinguish it among other figures in method section
* Modified the color theme of the methods section
* Refined the first paragraph of abstract to improve logical flow

1. **Poster presentation preparation and show case**
2. **Final Wrap-Up and future work discussion**

**Week 9**

1. **Produced a first draft of the poster**, which is mostly done.
2. **Communicated with Nicole to get feedback**. The following feedback points were noted and refined.

* Remove superscript on authors.
* Change affiliation to "Berkeley Design Summer Scholars."
* The first paragraph of the abstract needs to mention that consumer preferences can make sustainable design effective.
* The last paragraph of the abstract needs to remove repetitive technical details and briefly yet precisely state the effects achieved using NLP.
* Color the overview diagram in the methods section and adjust it for overall color themes.
* Remove BERT icons in the methods section.
* Add titles for the LLM selection chart.
* Add titles for textual results, including ABSA results and modeled topics.
* Reformat the textual failures and opportunities section to add emphasis.
* Reformat word formatting for findings and icons too.
* References need to include author, title, and year.
* Add a section for Amazon Sustainability Certification that explains abbreviations.

**Week 8**

1. **Database Creation and Chart Reorganization:** Created a new database to record and reorganize the charts produced from ABSA V2. These tables show customer sentiment towards each aspect and all certifications, including the percentage of positive and negative sentiments, as well as the total sentiment.
2. **Data Analysis and Extra Sheet Creation:** Created an extra sheet to modify the entries to display the positive to negative ratio, as well as the average percentage of positive and negative sentiment for both certified and non-certified features across all certifications.
3. **Summary of Findings:**

* Customers prioritize certain sustainability features more than others, with quality/performance being the most significant.
* Certified sustainability features, such as bio-friendliness and energy consumption, generally show a higher positive to negative sentiment ratio in customer reviews.
* Some certifications, such as Fair For Life, FSC, and RCS 100, show higher positive sentiment for aspects specific to their focus areas.
* Anomalies were found with Blue Angel certified products, which received more negative sentiment compared to other certifications, especially in general sustainability aspects.

1. **Created an overview diagram of the entire research approach.**
2. **Next Steps:**

* Conclude specific and organized design opportunities and failures using possibly charts and other tools.
* Poster making (ppt sized, title, berkeley logo)
  + Background
  + Methods (Overview Diagram, Example Images or annotation of high level output for each step)
  + High Level Results (icons for qualitative, and bar charts for quantitative)
  + Future Works (expanding on product description)

**Week 7**

1. **Program Fixes and Enhancements:** Fixed the program for ABSA V2, ensuring it produces accurate results for each certification. The output now includes organized Excel files, images, and modeled topics categorized by aspects and sentiment. Each aspect has separate positive and negative sentiment Excel files, as well as corresponding modeled topics.
2. **Running and Analyzing Results:** Ran the program to produce results for each certification. Following the analysis method decided last week, I have identified the opportunities and failures derived from the corresponding aspects covered by each certification. Additionally, for each certification, I have created a Word document specifying the certification, its covered range of products, and the identified opportunities and failures.
3. **Next Steps:** Refinement and Cross-Certification Analysis:

* Further refine the identified opportunities and failures by organizing them more effectively.
* Conduct a cross-certification analysis to compare how customers perceive sustainability features across different certifications.
* Overview diagram of (thought) process
* Quantitative graphs for cross-certification

**Week 6**

1. **Sustainability Keywords Modification:** Modified the list of sustainability keywords by adding words to each existing category and introducing some new categories.
2. **Certification Description Inspection:** Noticed that Amazon’s description of certification is not accurate so decided to re-evaluate the certifications. Inspected the description of each sustainability certification and marked the corresponding categories in Excel for easier comparison later.
3. **Analysis Method Decision:** Decided on the analysis method:
   * For each certification, assess how customers prioritize the certified sustainability aspects by comparing the percentage of comments that mention each aspect.
   * Conduct an aspect-based sentiment analysis on comments that mention each aspect to classify sentiment as positive or negative.
   * Using topic modeling, extract key elements that customers appreciate from positive sentiment comments and identify specific areas for potential improvement from negative sentiment comments.
4. **Initial Application of Analysis:** Applied this method to Fair For Life certified products but have not yet applied it to the other certifications. The program still has bugs to be fixed to entirely comply to this method.
5. **Next Steps: Further Application and Trend Summarization:**
   * Apply the analysis method to all other certifications.
   * Summarize the trends observed from the analysis across all certifications.

**Week 5**

1. Wrote website-bio, i.e. introduction of me on co-design lab wesite.
2. **Sustainability-Related Keywords Development:** This week, I focused on the paper "Combining Life Cycle Assessment and Online Customer Reviews to Design More Sustainable Products – Case Study on a Printing Machine." I read additional papers mentioned in it to identify sustainability-related keywords for developing aspects to be analyzed. Additionally, Nicole contacted Michael, the author of the paper, who provided 163 specific keywords directly. Using both resources, I categorized the keywords and developed a comprehensive database.
3. **Aspect-Based Sentiment Analysis:** Using the categorized keywords, I performed aspect-based sentiment analysis on the five databases of certifications. I obtained both visual and Excel results, providing a detailed sentiment analysis of the customer reviews based on the identified aspects.
4. **Further NLP Processing Methods:** I read additional papers about other NLP processing methods provided by Nicole, specifically focusing on attribute-sentiment analysis and ACOSI. I produced a comparison document summarizing the NLP processes I could potentially use. I am now ready to discuss with Nicole the further application of these methods in our research.
5. **Next Steps:** Currently, all five certifications are being analyzed using the same set of keywords. By tailoring the keywords and aspects for each certification, I aim to derive more specific insights that reflect the unique attributes of each certification. The next phase involves interpreting the results of the aspect-based sentiment analysis (ABSA) to determine what changes to the keywords are necessary for each certification to yield more meaningful and differentiated results. I have identified two approaches for this refinement:

* **Certification-Specific Aspects:** Add an aspect for each certification based on its description to observe user perception of certified products concerning the certification.
* **Topic Modeling Integration:** Incorporate aspects identified through topic modeling to refine the analysis further.

**Week 4**

1. **Combining Databases:** This week, I combined the reviews and product information databases for each of the five selected certifications. I introduced a unique key as a product ID for each item in the table and merged the data based on this ID. Now, I have five Excel databases, each corresponding to a certification and containing both reviews and product information for easier analysis.
2. **Sentiment Analysis:** I performed VADER and BERT sentiment analysis on each of the five certifications using Python programs after reading another paper “VADER vs. BERT A Comparative Performance Analysis for Sentiment on Coronavirus Outbreak”. I modeled the sentiment scores against customer ratings to evaluate the quality of the sentiment analysis. I am pleased with the BERT model, which produces sentiment scores aligning well with the ratings. These sentiment scores are now included in the databases, and I can show you the visual evaluation of the model next week.
3. **Quantitative and Qualitative Analysis Approach:** I applied quantitative sentiment analysis first because it is the final step of aspect-based sentiment analysis (qualitative). By grounding our approach in quantitative data, I ensure that our qualitative insights are well-supported. Here is a detailed explanation of my research approach:

* **Use Topic Modeling to Discover Themes:** Apply topic modeling to identify broad themes or topics within customer reviews. This helps in understanding the general areas of interest or concern among customers.
* **Refine Topics into Aspects:** Refine the discovered topics into more specific aspects of the product. This might involve manual interpretation or further processing to narrow down the topics into precise aspects.
* **Perform Sentiment Analysis on Aspects:** Apply sentiment analysis to each identified aspect to determine the sentiment expressed towards it. This can involve using pre-trained models or fine-tuning models on a labeled dataset.

1. **Next Steps:** Next week, I will proceed with topic modeling to discover themes and refine these into specific aspects for each certification or product, depending on the intermediate results. For future steps, I might use different domain-based knowledge to modify topic modeling and aspect extraction for each certification or product. This will help in tailoring our analysis to specific contexts and improving the accuracy of our insights.

**Week 3**

1. **Presentation Refinement:** Following Nicole's feedback from last week, I have refined the presentation. I also updated the presentation about my current progress since I have made extra progress since last week. Due to an unexpected event last week, the presentation has been rescheduled for next week.
2. **Product Scraper Progress:** The product scraper now generates a CSV file containing the following categories: product name, product image, price, product dimension, product description, sustainability features, rating, number of reviews, and product link. This deviates slightly from our original plan as it does not scrape material and weight information. This is because not all products list these details, making them technically challenging to scrape. However, material information can often be found within the product description.
3. **Review Scraper Issues:** The review scraper currently has a bug causing it to scrape only the first page of reviews multiple times, even when instructed to scrape multiple pages. I am actively working on fixing this issue.
4. **Redefining Certification Scope:** I have redefined the scope of the certification focus to concentrate on five certifications, each representing different aspects of sustainability. The goal is to infer design opportunities and failures from different sustainability perspectives, which will help compare insights later. I plan to scrape approximately 100 products for each certification and possibly 100 reviews for each product, totaling 10,000 reviews for each of the five certifications. I will assess the technical feasibility and adjust as necessary.

* **Blue Angel (Impacts of Product)**
* **BIFMA (Impacts during Manufacturing)**
* **Fair For Life (Impacts of Supply Chain)**
* **FSC (Impacts of Raw Materials on Forest)**
* **Recycled Claim Standard 100 (Impacts of Recyclable Raw Materials)**

1. **Current Focus and Next Steps:** This week, my primary focus is on refining and fixing the databases, as this process often requires more time than initially writing the program. I plan to start exploring natural language processing next week and believe I am still on track with the research timeline.
2. **Next Steps:** My tasks for the rest of this week include manually recording 100 Amazon links for each certification to input into the program and fixing the review scraper.

**Week 2**

1. **Draft Presentation Preparation:** I have prepared a draft PowerPoint presentation titled "Introduction to Research," which will be presented next week. This preliminary version focuses on content structure rather than visual refinement.
2. **Presentation Summary:**

* **Introduction:** Overview of the research question: How can we use sustainability certification information and customer reviews to identify design opportunities and failures?
* **Background:** Importance of sustainability in product development and customer choice, discussion on greenwashing, and different approaches to evaluating sustainability.
* **Literature Review:** Summary of three key papers on using machine learning, NLP, and LCA to extract sustainability insights from customer reviews. Insights on ML and NLP techniques for sustainable design, combining LCA with online reviews, and using LLMs for inferring product sustainability.
* **Methodology:** Steps to scrape Amazon products with sustainability certifications, process to scrape reviews for these products, and use NLP to gather sustainability-related insights, classifying insights into design failures and opportunities.
* **Current Progress:** Ongoing work on scraping Amazon products and reviews, initial findings, and next steps.

1. **Feedback and Revisions:** Nicole provided the following feedback:

* Add a slide introducing myself, including my name, major, hobbies, interests, and my collaborators (Nicole and Kosa).
* Modify slide 2 to replace "professional" with "grounded in data" and "unprofessional" with "potentially biased or not environmentally-informed".
* End the presentation with a slide outlining the next steps for the summer and ask the audience for feedback.
* Improve the visual style of the slides and ensure the presentation fits within a 10-minute timeframe.

1. **Database Exploration and Custom Scrapers:** I have been exploring databases of Amazon products and reviews, focusing on those with sustainability features. However, these databases did not align well with my research needs or were not comprehensive enough. As a result, I developed custom scrapers using open-source code for educational purposes. The scrapers for both Amazon products and reviews are currently about 80% functional, though some technical details still need refining.
2. **Next Steps:** I plan to refine the scrapers further and explore how to use NLP and LLM techniques to process the reviews and certifications from my scraped databases. I believe it would be more efficient to demonstrate these scrapers in person during our next meeting.

**Week 1**

1. **Program Kick-Off and Logistic Overview:** The program officially commenced with a comprehensive kick-off meeting and logistic overview. All necessary administrative and operational details were discussed to ensure a smooth start.
2. **Familiarization with Research Topic and Task:** I have thoroughly familiarized myself with the research topic and the specific tasks involved. This initial phase has provided me with a clear understanding of the project's objectives and my role within it.
3. **Literature Review and Research Preparation:** I have read three papers provided by Nicole that are pertinent to the research topic. Additionally, I reviewed the "Introduction to Research" to further solidify my foundational knowledge.
4. **Creation of Literature Review Document:** I created a detailed literature review which includes:

* **Summarization:** A concise summary of each paper.
* **Findings:** Key findings from the literature.
* **Relevance to Research Topic:** Analysis of how each paper contributes to our understanding of the research topic.
* **Limitations:** Identification of the limitations within the existing literature.

1. **Document Compilation:** All the above elements have been compiled into a comprehensive "Introduction to Research" document in Word format. This document will serve as a foundational reference for ongoing research activities.