ENVIRONMENTAL IMPACT OF THE CLOTHING INDUSTRY

GITHUB LINK: https://github.com/janhavi0210/clothing_industry

Information About the Dataset

The dataset has been sourced from Kaggle and is titled “PLASTIC-BASED TEXTILES IN CLOTHING INDUSTRY”. The dataset includes the amount of greenhouse gases & pollutants emitted, water & energy consumed, waste generated and sales revenue of well-renowned fashion brands: Zara, Nike, Urban Outfitters, Adidas and Forever 21 from the years 2018 - 2022. A caveat to note here is that there are no units mentioned for the numeric fields, so the analysis is based on the differences of magnitudes observed, not the magnitudes themselves.

Link: https://www.kaggle.com/datasets/purohitgaurav/plastic-based-textiles-in-clothing-industry

These brands are unarguably a few of the most consumed brands on the planet, especially amongst the middle to upper middle income class. These brands are mass producers, and this practice has grown uncontrollably recently - not just amongst these brands, but as a whole on the global level. Another important point to note is that in order to maximize production and sales, companies often resort to utilizing plastic based textiles such as Nylon, Polyester, and more. As a result of the high consumption of these brands, it is necessary to understand the level of impact these production practices have on the environment.

The potential insights this project aims to explore from this dataset are:

Revenue Growth of Companies over the years: Which company is ahead of the race?	Revenue vs. Greenhouse Gas Emissions: Is their a relationship between high revenue and high gas emissions?	Relation Between the Material and Environmental Impacts: Is there a material that is more sustainable than the others?

Data Preprocessing

The cleaned dataset is titled “Cleaned Dataset” and is a csv file. The python script has been executed in Jupyter Notebook. It is title “Data Preprocessing” and is a python notebook (ipynb) file. Both the datasets and the script are in the e-mail and can be found in the github repository as well.

The script contains the decisions as comments. Overall, the following preprocessing was done:

Finding and ensuring lack of null values & duplicates	Ensuring that no company is over or underrepresented	Renaming columns and product type to make it more visualization friendly	Making subsets of data to be compatible with Flourish

Data Visualization

In the previous step, two additional datasets were created to enable easier visualization in flourish. The datasets titled ‘Flourish Viz 1 and 2.csv’ and ‘Flourish Viz 3.csv’ are in the repository as well.

The visualizations are:

1. Revenue over the years: https://public.flourish.studio/visualisation/17366380/
- Main takeaway: Urban outfitters has become the brand with the highest sales revenue, Forever 21 being the lowest
2. Relation between Revenue and Greenhouse gas emissions: https://public.flourish.studio/visualisation/17366449/
- Main takeaway: Urban outfitters with the highest revenue in 2022, also has the highest Greenhouse Gas Emissions! **(The time slider needs to be played)**
3. Relation between product type and Greenhouse gas emissions: https://public.flourish.studio/visualisation/17366715/
- Main takeaway: Surprisingly, all types of materials have similar amounts of greenhouse gas production, some natural fibres being on top of the list! It might be attributed to demand of these materials and production practices.

Presentation

Link:

https://www.canva.com/design/DAGBCqaOTuU/fL3qta6f4g-HYxZINZ2cNA/view?utm_content=DAGBCqaOTuU&utm_campaign=designshare&utm_medium=link&utm_source=ed

itor

Note: The visualization on Slide 7 is a screen recording of the “Relation between product type and Greenhouse gas emissions” visualization, since it has a time slider.