Sustainability in Fashion and its Global Impacts

December 2, 2025

Olivia Newkirk

CSE 163

Research Questions

1. Which areas of the world have access to the most sustainable fashion and how does it reflect the surrounding environment?

The countries that have the most access to sustainable fashion are Italy, Japan, and India. All three of these countries feature diverse ecosystems. Germany, the UK and Brazil have the lowest access to sustainable fashion.

2. What production methods are considered the most sustainable, and how widely are they used globally?

Water usage can be considered to be slightly more sustainable than other production method practices, although I wouldn't consider any production methods to be sustainable. All methods are used very widely across the world.

3. Has sustainability in fashion improved throughout the years or gotten worse, and what factors are responsible for the change?

Sustainability in fashion depends on whether or not the market is growing, stable, or declining. For growing markets, sustainability has decreased overall. For stable markets, sustainability has increased overall. And finally for declining markets, sustainability has had no change.

Motivation

As an avid fashion lover and someone who enjoys thrifting and secondhand shopping, I'm very aware of the negative effects of fast fashion, and I want more insight on how exactly fast fashion impacts our environment. The sustainable fashion movement has seemed to quickly gain popularity after the general public was made aware of the horrors that go on in the fast fashion industry, with multiple factors contributing to the crisis. Child labor, extremely quick trend cycles, harmful chemicals released in the environment, and clothing ending up in the landfill are just a few of the byproducts that the fast fashion industry has and continues to produce. As of 2025, the fast fashion industry is worth approximately 150 billion. In comparison, the sustainable fashion market is only worth about 10 billion. That's a 1500% difference between industries. Although fast fashion is huge in the world right now, I want to understand what exactly is the driving force behind either sustainability or wasteful fashion.

Data Setting

https://www.kaggle.com/datasets/waqi786/sustainable-fashion-eco-friendly-trends/data

The dataset that I will use is from Kaggle. It features a list of 5000 fashion brands and various variables that are associated with sustainability. The dataset explores the country in which the brand is located, the brand ID, which is an anonymous way to categorize the brand to avoid legal issues, the percentage of sustainable materials that the brand uses, the production process for producing the clothing, and the year that the data was collected.

As for the context of the dataset and how it could complicate my analysis, the geographical differences and cultural biases could influence how sustainability is structurally different in other countries, and it may be a confounding variable in some ways. The vagueness of some definitions like "sustainability score" isn't clearly defined so it could be difficult to provide a clear definition of what goes into determining the sustainability of the brand. The third potential complication could be the anonymity of brand name, which could make it difficult to pinpoint exactly which brands are less sustainable than others, which would be helpful from a consumer standpoint.

Method

1. Load and clean the dataset

   This makes sure the data is clean and ready to analyze.

   ```
   with open("fashion.csv") as f:
   clean_data("fashion.csv") - data manipulation
   ```

   Test on the fashion dataset

2. Compute summaries

   This will help answer the research question "Which areas of the world have access to the most sustainable fashion and how does it reflect the surrounding environment?" and information about production methods.

   ```
   country_summary() - data manipulation
    Test on sample of the fashion dataset
   ```

3. Independent T Test

   This will support my hypothesis which will help me achieve my challenge goal of statistical hypothesis testing.

   ```
   correlation() - data manipulation
   Test on data set
   ```

EDA Results

1. I targeted each research question and I analyzed my questions one at a time. Tackling each question one by one made it a lot easier to work through the project and make it less overwhelming.

2. I learned that my data was already pretty clean, and I didn't have to spend a lot of time cleaning the data. Overall, the data analysis process was fun and relatively simple.

3. My understanding of sustainability scores increased, as I was orginally confused on what that metric meant. After working through the data, I believe it means all the production methods as a mean

[50]:
```python
import pandas as pd
import seaborn as sns
import io
import matplotlib.pyplot as plt
import doctest
sns.set_theme()
```

[51]:
```python
# Altair importation
import altair as alt
```

[79]:
```python
# Testing dataset #1

testing = pd.DataFrame({
    "Eco_Friendly_Manufacturing": ["yes", "yes", "no", "no", "yes", "no"],
    "Waste_Production_KG": [120, 150, 400, 380, 130, 420]
})
```

[82]:
```python
# Testing dataset #2

testing2 = pd.DataFrame({
    "Eco_Friendly_Manufacturing": ["yes", "yes", "no", "no", "yes", "no"],
    "Carbon_Footprint_MT": [10, 12, 25, 22, 11, 24],
    "Water_Usage_Liters": [1000, 1100, 3000, 2800, 900, 3100],
    "Waste_Production_KG": [120, 150, 400, 380, 130, 420]
})
```

[52]:
```python
fashion = pd.read_csv("fashion.csv")
fashion
```

[52]:
```
        Brand_ID  Brand_Name    Country  Year Sustainability_Rating  \
0     BRAND-0001     Brand_1  Australia  2018                     D
1     BRAND-0002     Brand_2      Japan  2015                     D
2     BRAND-0003     Brand_3        USA  2024                     A
3     BRAND-0004     Brand_4      Italy  2023                     D
4     BRAND-0005     Brand_5        USA  2016                     D
…            …           …          …     …                    …
4995  BRAND-4996  Brand_4996     Brazil  2010                     A
4996  BRAND-4997  Brand_4997    Germany  2012                     C
4997  BRAND-4998  Brand_4998      Japan  2023                     D
4998  BRAND-4999  Brand_4999      India  2016                     A
4999  BRAND-5000  Brand_5000      Italy  2016                     B

       Material_Type Eco_Friendly_Manufacturing  Carbon_Footprint_MT  \
0             Tencel                         No                 1.75
1       Vegan Leather                       Yes               124.39
2       Vegan Leather                        No               336.66
3       Bamboo Fabric                        No               152.04
```

```
4      Bamboo Fabric                          Yes              415.63
…             …                        …                  …
4995  Organic Cotton                          Yes              423.73
4996   Bamboo Fabric                          Yes               68.24
4997   Vegan Leather                          Yes              321.12
4998          Tencel                           No              124.83
4999   Vegan Leather                           No              192.11

      Water_Usage_Liters  Waste_Production_KG Recycling_Programs  \
0             4511152.79             97844.11                 No
1             1951566.31             37267.75                 No
2              467454.52             38385.92                 No
3              899576.90             32665.45                 No
4             1809219.90             37295.47                Yes
…                    …                 …                 …
4995           251736.61             50532.28                 No
4996          2342347.78             64214.57                Yes
4997           317800.67             81467.21                Yes
4998           927539.68             34506.22                Yes
4999          3402552.70             81958.82                Yes

      Product_Lines  Average_Price_USD Market_Trend Certifications
0                 2              38.33      Growing           GOTS
1                15             250.07      Growing           GOTS
2                 2             146.16      Growing         B Corp
3                13             165.52       Stable       OEKO-TEX
4                19             211.63       Stable     Fair Trade
…             …                 …           …            …
4995              1             490.95       Stable            NaN
4996             13              53.92       Stable     Fair Trade
4997             12             431.32       Stable            NaN
4998             20             322.95    Declining     Fair Trade
4999             11             205.62      Growing            NaN

[5000 rows x 15 columns]
```

There are 5000 rows and 15 columns in the dataset. The rows represent a singular brand and the columns represent different features of brand, including the brand ID, brand name, country, year, sustainability rating, material, eco friendly manufacturing process, carbon footprint, water usage, water production, recycling programs, product lines, average price of an item, market trend, and certifications.

Does the dataset have any missing values?

[53]:
```
missing_values = fashion.isna().sum()
missing_values
```

[53]:
```
Brand_ID                       0
Brand_Name                     0
Country                        0
Year                           0
Sustainability_Rating          0
Material_Type                  0
Eco_Friendly_Manufacturing     0
Carbon_Footprint_MT            0
Water_Usage_Liters             0
Waste_Production_KG            0
Recycling_Programs             0
Product_Lines                  0
Average_Price_USD              0
Market_Trend                   0
Certifications              1004
dtype: int64
```

The dataset does have missing data in the Certifications columns, and my plan for working around this is to not use the Certifications column as it doesn't pertain to my research questions.

Variables of Interest

My variables of interest are the country, year, sustainability rating, eco friendly manufacturing, carbon footprint, waste production, recycling programs, and market trend.

The brand name represents each individual brand. The country represents the country in which the brand is based in. The year represents the year in which the sustainability measurements for the individual brand have taken place. The sustainability rating represents the overall sustainability based on the sustainability related variables in the dataset. Eco friendly manufacturing represents whether or not there is an overall amount of eco-friendly manufactoring practices that the brand uses. Carbon footprint represents the total amount of carbon dioxide in metric tons that were emitted by the brand in the given year. Waste production represents the amount of waste in kilograms produced by the brand in a given year. Recycling programs represents any recycling program that the brand is or has participating in the given year. Market trend represents whether or not the brand is growing, stable, or declining.

As for my research questions, the variables that I am matching to my first research question, "Which areas of the world have access to the most sustainable fashion and how does it reflect the surrounding environment?", are "country", "year", "sustainability rating", "eco-friendly manufacturing", "carbon footprint", "waste production", and "recycling programs".

The variables I will be using for the second research question, "What production methods are considered the most sustainable, and how widely are they used globally?", are "country", "year", "sustainability rating", "eco-friendly manufacturing", "carbon footprint", "waste production", and "recycling programs".

The variables I will be using for the third research question, "Has sustainability in fashion improved throughout the years or gotten worse, and what factors are responsible for the change?", are "year", "sustainability rating", "eco friendly manufacturing", "carbon footprint", "waste production", "recycling programs", and "market trend".

I'm using most of the columns in the dataset because I appreciate how the dataset has been a general sustainability rating, and multiple factors of what goes into the rating. Using all of the production related factors will allow my analysis to be deeper and more comprehensive.

Variable Summaries

[54]:
```python
fashion['Year'].describe()
```

[54]:
```
count    5000.000000
mean     2016.956200
std         4.347466
min      2010.000000
25%      2013.000000
50%      2017.000000
75%      2021.000000
max      2024.000000
Name: Year, dtype: float64
```

[55]:
```python
fashion['Country'].describe()
```

[55]:
```
count      5000
unique       10
top          UK
freq        515
Name: Country, dtype: object
```

[56]:
```python
fashion['Sustainability_Rating'].describe()
```

[56]:
```
count      5000
unique        4
top           B
freq       1278
Name: Sustainability_Rating, dtype: object
```

[57]:
```python
fashion['Eco_Friendly_Manufacturing'].describe()
```

[57]:
```
count      5000
unique        2
top         Yes
freq       2520
Name: Eco_Friendly_Manufacturing, dtype: object
```

[58]:
```python
fashion['Carbon_Footprint_MT'].describe()
```

[58]:
```
count    5000.000000
mean      250.317550
std       142.802345
min         1.040000
```

```
25%        126.615000
50%        250.650000
75%        372.255000
max        499.930000
Name: Carbon_Footprint_MT, dtype: float64
```

[59]:
```
fashion['Water_Usage_Liters'].describe()
```

[59]:
```
count    5.000000e+03
mean     2.517862e+06
std      1.429043e+06
min      5.010618e+04
25%      1.293087e+06
50%      2.499096e+06
75%      3.763860e+06
max      4.999597e+06
Name: Water_Usage_Liters, dtype: float64
```

[60]:
```
fashion['Waste_Production_KG'].describe()
```

[60]:
```
count     5000.000000
mean     50107.304048
std      28746.381115
min       1026.380000
25%      25340.802500
50%      50466.340000
75%      74984.650000
max      99947.850000
Name: Waste_Production_KG, dtype: float64
```

[61]:
```
fashion['Recycling_Programs'].describe()
```

[61]:
```
count     5000
unique       2
top        Yes
freq      2528
Name: Recycling_Programs, dtype: object
```

[62]:
```
fashion['Market_Trend'].describe()
```

[62]:
```
count       5000
unique         3
top       Stable
freq        1709
Name: Market_Trend, dtype: object
```

Compare Production Methods and Byproducts

[87]:
```python
def comparison(fashion):
    """
    This function takes a fashion dataset and compares three production methods.
    It returns a dataframe containing the mean and the difference of each␣
↪method for
    eco friendly brands and non eco friendly brands.
    >>> comparison(fashion)
    >>> fashion['Eco_Friendly_Mean'].round(3).tolist()
    [2514177.0, 50485.9, 249.797]
    >>> fashion['Non_Eco_Mean'].round(3).tolist()
    [2521605.0, 49722.6, 250.846]
    >>> fashion['Difference'].round(3).tolist()
    [-7427.953, 763.300, -1.049]

    >>> result = comparison(testing2)
    >>> result.loc["Carbon_Footprint_MT", "Eco_Friendly_Mean"]
    11.0
    >>> result.loc["Carbon_Footprint_MT", "Non_Eco_Mean"]
    23.666666666666668
    >>> result.loc["Water_Usage_Liters", "Difference"]
    -1966.6666666666665
    >>> result.loc["Waste_Production_KG", "Eco_Friendly_Mean"]
    133.33333333333334
    >>> result.loc["Waste_Production_KG", "Non_Eco_Mean"]
    400.0
    """
    fashion['eco'] = fashion['Eco_Friendly_Manufacturing'].str.lower() == 'yes'

    production_methods = [
        'Carbon_Footprint_MT',
        'Water_Usage_Liters',
        'Waste_Production_KG']

    eco_means = fashion[fashion['eco'] == True][production_methods].mean()
    non_eco_means = fashion[fashion['eco'] == False][production_methods].mean()

    comparison = pd.DataFrame({
        'Eco_Friendly_Mean': eco_means,
        'Non_Eco_Mean': non_eco_means
    })

    comparison['Difference'] = comparison['Eco_Friendly_Mean'] -␣
↪comparison['Non_Eco_Mean']
    comparison = comparison.sort_values('Difference', key=abs, ascending=False)

    print(comparison)
```

```
comparison(fashion)
```

```
                    Eco_Friendly_Mean  Non_Eco_Mean   Difference
Water_Usage_Liters       2.514177e+06  2.521605e+06 -7427.953096
Waste_Production_KG      5.048590e+04  4.972260e+04   763.300418
Carbon_Footprint_MT      2.497971e+02  2.508464e+02    -1.049328
```

Results

This analysis addresses the research question "What production methods are considered the most sustainable, and how widely are they used globally?". Eco friendly brands use slightly less liters of water while producing clothing on average than non-eco friendly brands. Eco friendly brands produce slightly more waste than non-eco friendly brands. Finally, Eco friendly brands produce slightly less carbon footprint than non eco friendly brands. I find these slight differences extremely interested and unexpected, I would have thought that brands that are considered eco friendly would not contribute this much to harmful production practices. This could suggest that clothing production in general is not kind to the planet, and there's little to no way around mass production not being environmentally harmful.

Dataset Visualizations

[90]:
```
rating_counts = (fashion.groupby(['Country', 'Sustainability_Rating']).size().
    ↪reset_index(name = 'Count'))
top_rating = (rating_counts.loc[rating_counts.groupby('Country')['Count'].
    ↪idxmax()])
print(top_rating)
```

```
      Country Sustainability_Rating  Count
2   Australia                     C    133
5      Brazil                     B    135
11      China                     D    138
12     France                     A    132
17    Germany                     B    136
20      India                     A    134
24      Italy                     A    139
28      Japan                     A    125
34         UK                     C    141
39        USA                     D    138
```

[91]:
```
a_ratings = (
    fashion.assign(A_Rating = fashion['Sustainability_Rating'] == 'A')
    .groupby('Country', as_index=False)['A_Rating']

    .mean()
)

a_ratings = a_ratings.sort_values('A_Rating', ascending=False)
a_ratings['Country'] = a_ratings['Country'].astype(str)

```

```
chart = (
    alt.Chart(a_ratings)
    .mark_bar()
    .encode(
        x=alt.X('Country:N', sort='-y', axis=alt.Axis(labelAngle=-45)),
        y=alt.Y('A_Rating:Q', title='"A" Rating Percentage'),
        color=alt.Color('A_Rating:Q', scale=alt.Scale(scheme='greens')),
        tooltip=['Country', 'A_Rating']
    )
    .properties(
        width=600,
        height=400,
        title='Percentage of "A" Sustainability Ratings per Country'
    )
)

chart
```

[91]: alt.Chart(…)

This bar chart shows the percentage of fashion brands in each country who have received an A score in overall sustainability.

A vertical bar chart displaying ten countries on the x-axis and the percentage of "A" ratings on the y-axis. Countries with taller green bars represent a higher proportion of sustainable brands.

This plot compares the ratings of the top sustainable fashion brands across countries. A bar chart is used because both variables are categorical, and it clearly displays differences between countries. Viewers can interpret which countries lead in sustainable fashion practices through the color differences and the heights of the bars.

Results

This plot and analysis addresses the research question "Which areas of the world have access to the most sustainable fashion and how does it reflect the surrounding environment?". I found the top three countries that had the highest proportion of A sustainability rated brands were Italy, Japan, and India. The bottom three countries that had the lowest proportion of A sustainability rated brands were Germany, the UK, and Brazil. I find this very interesting because Italy, Japan, and India are all geographically pretty different from each other. I initially wondered if I would find that the most sustainable brands would be grouped together by continent, but it appears they are not.

[92]:
```
chart2 = (
    alt.Chart(fashion)
    .mark_bar()
    .encode(
        x=alt.X('Country:N', sort='-y', axis=alt.Axis(labelAngle=-45)),
        y=alt.Y('Carbon_Footprint_MT:Q', title='Carbon Footprint in Metric␣
 ↪Tons'),
```

```
        color=alt.Color('Carbon_Footprint_MT:Q', scale=alt.
↪Scale(scheme='oranges')),
        tooltip=['Country', 'Carbon_Footprint_MT']
    )
    .properties(
        width=600,
        height=400,
        title='Country vs Carbon Footprint'
    )
)

chart2
```

[92]: alt.Chart(…)

This bar chart shows the average carbon footprint in each country.

A vertical bar chart displaying ten countries on the x-axis and carbon footprint on the y-axis. Countries with taller orange bars represent a higher amount of carbon emissions.

This plot uses average carbon emissions to visualize negative environmental impact by country. It's appropriate because carbon footprint is a continuous variable, and bar charts help compare the impact across categorical variables. Readers can quickly spot countries with larger carbon footprint.

Results

This plot addresses the research question, "Which areas of the world have access to the most sustainable fashion and how does it reflect the surrounding environment?" I found that Japan had the lowest carbon footprint compared to the other countries, with a considerable dip in the visualization. That makes sense, because Japan is one of the three countries with the highest amount of "A" rated sustainability brands. Germany and France are among countries that have a lower carbon footprint as well, which is interesting because they are both European countries. Something that I have to consider with carbon footprint is the size of the countries, which Japan, Germany, and France are all small countries compared to the other ones listed (exceptions to Italy and the UK). In a further analysis, I would highly consider taking country size into account.

[93]:
```
chart3 = (
    alt.Chart(fashion)
    .mark_bar()
    .encode(
        x=alt.X(
            'Eco_Friendly_Manufacturing:N',
            title='Eco Friendly Manufacturing',
            sort=['No', 'Yes']
        ),
        y=alt.Y(
            'Waste_Production_KG:Q',
            title='Waste Production (KG)'
```

```
        ),
        color=alt.Color(
            'Eco_Friendly_Manufacturing:N',
            scale=alt.Scale(scheme='greens'),
            legend=None
        ),
        tooltip=['Eco_Friendly_Manufacturing', 'Waste_Production_KG']
    )
    .properties(
        width=300,
        height=350,
        title='Relationship Between Waste Production and Eco Friendly␣
 ↪Manufacturing'
    )
)

chart3
```

[93]: alt.Chart(…)

This bar chart compares average waste production for brands that use eco-friendly manufacturing versus those that do not.

Two green bars compare average waste (in kilograms) between eco-friendly and non-eco-friendly brands.

This visualization directly compares production methods. It shows whether eco-friendly manufacturing leads to less waste generation. Bar plots are effective here because they make differences in averages between categories clear.

Results

This plot addresses the research question "What production methods are considered the most sustainable, and how widely are they used globally?". I found that waste production between eco friendly and non eco friendly brands is barely noticable, which was confirmed by my t test I performed below. I find this very interesting and unexpected because I naturally would have thought that eco friendly brands would produce significantly less waste, which is not the case. This tells me that I need to look into the criteria that seperates an eco friendly brand from a non eco friendly brands, as it's unclear, especially with this result. However, this may mean that even though a brand is considered eco friendly, waste production is inevitable.

[94]:
```
chart4 = (
    alt.Chart(fashion.dropna(subset=['Carbon_Footprint_MT', 'Year']))
    .mark_line(point=True)
    .encode(
        x=alt.X('Year:O', title='Year'),
        y=alt.Y(
            'mean(Carbon_Footprint_MT):Q',
            title='Mean Carbon Footprint (MT)',
```

```
            scale=alt.Scale(domain=[220, 280])
        ),
        color=alt.Color('Market_Trend:N', title='Market Trend', scale=alt.
↪Scale(scheme='set2')),
        tooltip=[
            alt.Tooltip('Year:O'),
            alt.Tooltip('Market_Trend:N', title='Trend'),
            alt.Tooltip('mean(Carbon_Footprint_MT):Q', format=',.0f',␣
↪title='Avg Footprint')
        ]
    )
    .properties(
        width=550,
        height=350,
        title='Carbon Footprint by Year and Market Trend'
    )
)

chart4
```

[94]: alt.Chart(…)

This line plot tracks how average carbon footprint changes over time, with separate lines for growing, stable, and declining market trends.

A line plot shows carbon emissions by year, with separate colored lines for market trends. Each line rises or falls depending on whether sustainability is improving throughout the years.

By plotting average carbon footprint over time, this visualization reveals whether sustainability efforts are improving, declining, or staying the same. This shows readers how different markets have an effect on sustainability.

Results

This plot addresses the research question "Has sustainability in fashion improved throughout the years or gotten worse, and what factors are responsible for the change?". Overall, all market trends have experienced levels of volatility throughout the years, but some carbon footprints have improved and some have gotten worse. To clarify, if carbon footprint declines, that's considered an improvement, and if carbon footprint increases, that's considered a decline. The market trend that experienced overall improvement was "stable". The market trend that experienced overall decline was "growing". And finally, the market trend that experienced a small change was "declining". These results are the most interesting to unpack in my opinion. As brands are increasing in popularity, carbon footprint subsequently increases with them. This result makes a lot of sense, and lines up with my expectations. What's particularily interesting is carbon footprint decreasing for stable markets, and staying the same for declining markets. I would have expected carbon footprint to stay the same for stable markets and decrease for declining markets, so I'm curious into looking more into this in the future.

Result Validity

Research Question 2: What production methods are considered the most sustainable, and how widely are they used globally?

Null hypothesis: There is no difference in waste production between eco-friendly and non-eco friendly brands.

Alternative hypothesis: Brands using eco-friendly manufacturing produce less waste on average.

Statistical test: Independent sampled t test

Assumptions: Samples are independent from each other, normally distributed, and equal variances.

[96]:
```python
def t_test(fashion):
    """
    This function takes a fashion dataset and returns a summary of indepedent␣
↪sampled t test
    results for eco friendly and non eco friendly, including means of both␣
↪variables,
    variances of both variables, pooled variance,
    standard error, t statistic, and degrees of freedom.

    >>> t_test(fashion)
    Eco mean: 50485.90105555552
    Non-eco mean: 49722.600637096875
    Eco variance: 809231083.4534745
    Non-eco variance: 843793644.4555906
    Pooled variance: 826374058.3882976
    Standard error: 813.106118394434
    T statistic: 0.9387463717108234
    Degrees of freedom: 4998

    >>> t_test(testing)
    Eco mean: 133.33333333333334
    Non-eco mean: 400.0
    Eco variance: 233.33333333333337
    Non-eco variance: 400.0
    Pooled variance: 316.6666666666667
    Standard error: 14.529663145135558
    T statistic: -18.353258709644937
    Degrees of freedom: 4
    """
    fashion['eco_friendly'] = fashion['Eco_Friendly_Manufacturing'].str.lower()␣
↪== 'yes'

    eco = fashion[fashion['eco_friendly'] == True]['Waste_Production_KG']
    non_eco = fashion[fashion['eco_friendly'] == False]['Waste_Production_KG']

    mean1 = sum(eco) / len(eco)
    mean2 = sum(non_eco) / len(non_eco)
```

```
    sum1 = sum(eco)
    sum2 = sum(non_eco)

    sample1 = len(eco)
    sample2 = len(non_eco)

    diff1 = 0
    for x in eco:
        diff1 += (x - mean1) ** 2
    var1 = diff1 / (sample1 - 1)

    diff2 = 0
    for x in non_eco:
        diff2 += (x - mean2) ** 2
    var2 = diff2 / (sample2 - 1)

    import math
    pooled_var = ((sample1 - 1)*var1 + (sample2 - 1)*var2) / (sample1 + sample2␣
 ↪- 2)

    standard_error = math.sqrt(pooled_var * (1/sample1 + 1/sample2))

    t_statistic = (mean1 - mean2) / standard_error

    df = sample1 + sample2 - 2

    print("Eco mean:", mean1)
    print("Non-eco mean:", mean2)
    print("Eco variance:", var1)
    print("Non-eco variance:", var2)
    print("Pooled variance:", pooled_var)
    print("Standard error:", standard_error)
    print("T statistic:", t_statistic)
    print("Degrees of freedom:", df)

t_test(fashion)
doctest.run_docstring_examples(t_test, globals())
```

```
Eco mean: 50485.90105555552
Non-eco mean: 49722.600637096875
Eco variance: 809231083.4534745
Non-eco variance: 843793644.4555906
Pooled variance: 826374058.3882976
Standard error: 813.106118394434
T statistic: 0.9387463717108234
Degrees of freedom: 4998
```

T Test Interpretation

By subtracting the non-eco mean by the eco mean, I found out that eco friendly brands produce slightly more waste on average, which is unexpected. Overall, there is no statistical evidence that eco friendly brands produce less waste on average compared to non-eco friendly brands. This means we fail to reject the null hypothesis.

Impact and Limitations

1. I think that both consumers and other companies would find my results beneficial. My results reveal that are not significant production methods that define a brand as “sustainable” or “not sustainable”, which I was personally surprised by. Overall, I think my findings will help brands improve sustainability within themselves, and hopefully lead to insights that will help the planet.

2. The data setting categorizing brands as anonymous hurt my overall analysis because I wasn’t able to make any statements or analysis regarding specific brands. This would have provided consumers with brand transparency. Most fashion brands aren’t exactly known for transparency, so I’m not exactly surprised on having to use an anonymous dataset. Since I sourced the data from Kaggle, it’s hard to verify whether or not the data is completely accuarate. When I was in the process of selecting my dataset, I tried looking for official data but it was very difficult or locked behind a paywall.

3.

- I didn’t complete correlation coefficients or validity testing for all of my research questions, just question two. If I fleshed my analysis in this way, I believe my conclusions would be even more solid.
- My analysis doesn’t include every variable available in the dataset. I didn’t use recycling programs or product lines variables, and using them could’ve deepened my analysis.
- The dataset only contained from years 2010 through 2024. This limited the amount of analysis I could perform, since I was limited to a 14 year timespan. If I had access to more years, I could’ve performed more in depth analysis of how sustainability has changed over time.
- The lack of transparency about what constitutes an “eco friendly” brand versus a “non-eco friendly” brand.
- I also wish I was able to cross analyze this dataset with an global evironment dataset to further connect ecosystems and the impact that fast fashion has on them. I may do this in the future out of curiosity!

Challenge Goals

New Library:

I ran into a few challenges with the New Library challenge goal. I orginally planned to use plotly as my new library, but I tried to install and load the package and it failed to load the graphs in Juptyer Hub. Because of this drawback, I decided to use a different new libarary for my graphing, altair. Figuring out this library was definitely a learning curve, and I wish I was able to use plotly because it’s simpler to use and understand.

Hypothesis Testing:

I was able to conduct an independent two sampled t test with my second research question. This provided me with a little more insight on figuring out results for this research question. I selected a specific production method, waste production, to see if brands categorized as sustainable were less likely to produce more waste.

Plan Evaluation

Before Final Touches: Overall, I think this EDA took me a lot longer than I realized, and I think that I need to adjust the time spent on my project in general. The tasks that still need to be completed are new library implementation, refining ideas for my visualizations, and extra result validity.

After Final Touches: The time I alotted myself makes a lot of sense, and I was able to manage my time and figure out setbacks. I was able to implement both of my challenge goals, complete the 120 lines of code minimum, and write docstrings and doctests.

Testing

I created two functions to prep and perform analysis on my dataset, and I wrote two docstrings and one doctest for each function.