Project: Does Mental Health Affect your Appearance? - Data Analysis

Table of Content

Project Description

> The purpose of this project is to find if Mental Health has an effect on an individual's appearance. So, a survey was conducted to clean, analyze, and visualize the survey's data and come up with a conclusion answering the research question.

Research Question

Does Mental Health Affect your Appearance?

Hypothesis

An individual's mental health has a strong, positive effect on his/her appearance.

Population of Interest

This project is interested in people who have faced misfortunes in their lives that affected their mental health and appearance during that period.

Calculating the population's size is near impossible because many misfortunes occur in people's lives that may affect their appearance which makes it a difficult task to come up with a certain value for their population.

Sampling Method

Due to the limited access to research resources and population of interest, Convenience Sampling is used to help address the research question because ,as stated previously, the population is not easily accessible, thus, using the Simple Random Sampling method would make no sense as the population is not accurately defined.

From its definition, Convenience sampling involves using respondents who are “convenient” to the research question. There is no pattern whatsoever in acquiring these respondents—they may be recruited merely asking people who are present in the street, in a public building, or in a workplace, for example.

Bias Identification

There are potential sources of bias in the design and choice of questions in the sruvey, but bias cannot be eliminated for sure; it can be minimized.

Sources of bias:

1. At least 90% of the responses are within the 16-24 Age range (not accurately representative of the population)
2. At least 90% of the responses came from Egypt which makes the possible conclusions biased to Egyptians.

Steps to Minimize Sources of bias:

1. As stated previously, Convenience Sampling was used, so the survey was sent to accessible people within my network which are in the same age range.
2. The survey was sent to mental health & awareness groups on Facebook, threads on Reddit, and to friends over in the United States of America; responses were minimal in comparison to the responses from Egypt.

Survey Link

CLICK ME

Survey Snapshots

Does Mental Health Affect Your Appearance?

This form is for a Data Analysis University Project, so your response will be of much help. Data is anonymously collected for confidentiality.
The **purpose** of this form is to find out whether mental health can have an effect on your physical appearance.

Please take your time answering this form!

* Required

Age *

- ◯ 16-24
- ◯ 25-34
- ◯ 35-44
- ◯ 45+

Gender *

- ◯ Male
- ◯ Female
- ◯ Other: _____

Which country you live in? *

Your answer

Have you ever faced a situation that affected your mental health? *

- ◯ Yes
- ◯ No

On a scale of 1-5, how deeply was your mental health affected? *

1: Had no Effect

5: Deeply Affected

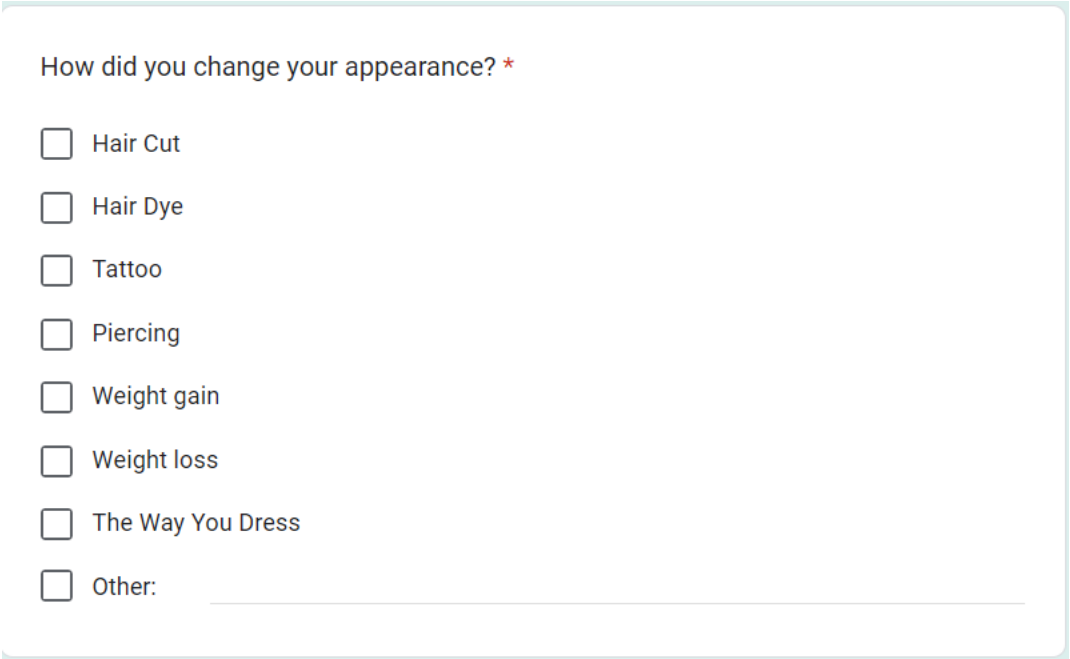
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5

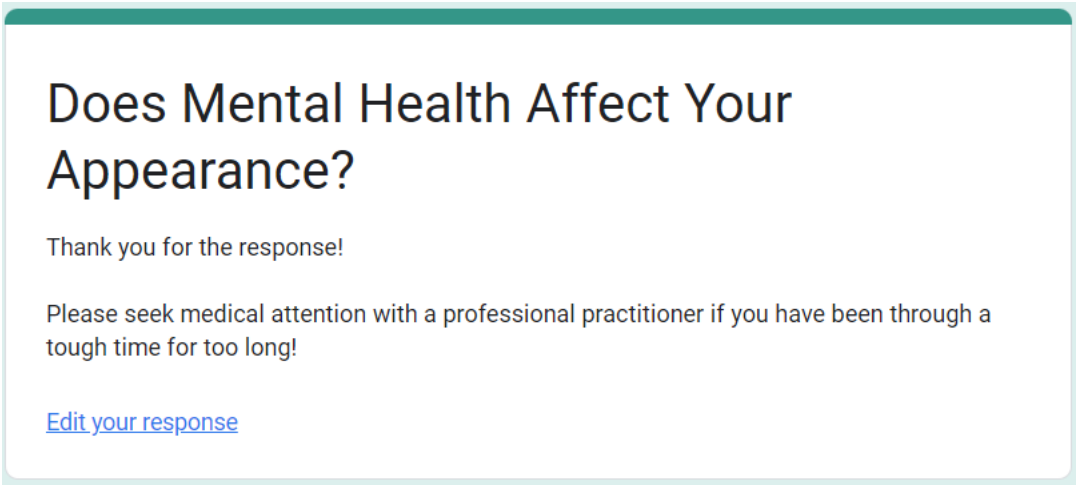
How long was your mental health affected? (in months)

- ◯ 0-3
- ◯ 3-6
- ◯ 6-9
- ◯ 9+

Have you changed your appearance during that time? *

- ◯ Yes
- ◯ No





Number of Samples & Responses Collected

1 sample

107 responses

Data Cleaning

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files
csv_file = files.upload()
pd.options.mode.chained_assignment = None  # default='warn'
```

Choose Files No file chosen

```
df = pd.read_csv("Does Mental Health Affect Your Appearance (Responses).csv")
```

```
df.head()
```

	Timestamp	Age	Gender	Which country you live in?	Have you ever faced a situation that affected your mental health?	On a scale of 1-5, how deeply was your mental health affected?	How long was your mental health affected? (in months)	Have you changed your appearance during that time?	How dic change appeara
0	3/5/2023 0:16:26	16-24	Female	Cairo	Yes	5	0-3	Yes	Pie Weigh
1	3/5/2023 18:00:52	16-24	Female	Cairo	Yes	4	0-3	Yes	The Wa
2	3/5/2023 18:04:16	16-24	Male	Cairo	Yes	5	3-6	No	

The naming convention of columns will be lower case, and two words or more will be seprated by an underscore "_".

The "Timestamp" column will be dropped as it is useless.

The questions asked will be reworded to short, concise words for readability purposes.

```
df.drop(["Timestamp"], axis=1, inplace= True)
new_column_names = ["age", "gender", "country", "mental_health_affected", "mental_health_scale", "mental_health_duration", "appearance_affect
df.columns = new_column_names
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 106 entries, 0 to 105
Data columns (total 8 columns):
 #   Column                  Non-Null Count  Dtype 
---  ------                  --------------  ----- 
 0   age                     106 non-null    object
 1   gender                  106 non-null    object
 2   country                 106 non-null    object
 3   mental_health_affected  106 non-null    object
 4   mental_health_scale     106 non-null    int64 
 5   mental_health_duration  91 non-null     object
 6   appearance_affected     91 non-null     object
 7   appearance_changed      61 non-null     object
dtypes: int64(1), object(7)
memory usage: 6.8+ KB
```

Next step is to change the dtype of columns that can be represented as booleans.

Disclaimer: In Pandas, 0 or 1 as booleans are mapped to False and True, and vice-versa.

Columns Applicable:

- mental_health_affected: 0 for No, 1 for Yes
- appearance_affected: 0 for No, 1 for Yes

```
df["mental_health_affected"] = df["mental_health_affected"].map({"Yes": True, "No": False})
df["appearance_affected"] = df["appearance_affected"].map({"Yes": True, "No": False})
```

```
df
```

	age	gender	country	mental_health_affected	mental_health_scale	mental_health_duration	appear
0	16-24	Female	Cairo	True	5	0-3	
1	16-24	Female	Cairo	True	4	0-3	
2	16-24	Male	Cairo	True	5	3-6	
3	16-24	Female	Cairo	False	1	NaN	
4	16-24	Female	Cairo	True	4	3-6	
...	...	...	...	...	...	...	
	16-		United				

Next step is to clean the "country" column input values as data input could not have been validated via Google Forms, change the dtype of the "mental_health_scale" from float64 to int64. Another error is that Excel auto-corrected the "mental_health_duration" column values to DateTime values.

In other words, "3-6" value was transformed to "6-Mar", and "6-9" to "9-Jun". These cases need to be handled too.

The "appearance_changed" column values will be used to calculate their count by splitting the string based on how many commas are there, and used later on to find if there is a relation between the "mental_health_scale" and the "appearance_changed" values.

```
df = df[df["mental_health_scale"].notna()]
df["mental_health_scale"] = df["mental_health_scale"].astype(int)
df["country"] = df["country"].str.lower()
country_conditions = [(df["country"].str.contains("egypt|cairo|damietta|eg|pt|giza|ismailia")), (df["country"].str.contains("u")), (df["count
countries = ["Egypt", "USA", "Scotland"]
df["country"]= np.select(country_conditions, countries)


print("Egypt: ", df['country'].value_counts()["Egypt"])
print("USA: ", df['country'].value_counts()["USA"])
print("Scotland: ", df['country'].value_counts()["Scotland"])
```

```
Egypt:  98
USA:  6
Scotland:  1
```

The sum of values counts up with the number of entries that mapped to these countries before cleaning, so mission successful!

```
df.dropna(subset=["mental_health_duration"])
df = df[df["gender"].isin(["Male","Female"])]
```

Since the population of interest is mentally affected individuals, any non-affected individuals' input data will be dropped.

```
df["mental_health_duration"] = df["mental_health_duration"].map({"6-Mar": "3-6", "9-Jun": "6-9", "0-3": "0-3", "9+": "9+"})
```

```
df["appearance_changed_count"] = df["appearance_changed"].str.count(',')
```

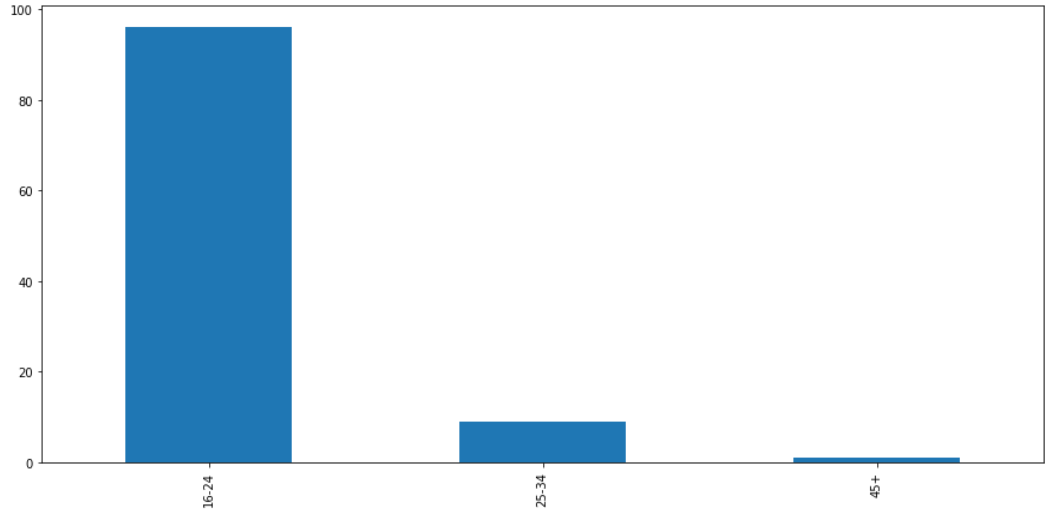
```
df.head(10)
```

	age	gender	country	mental_health_affected	mental_health_scale	mental_health_duration	appearanc
0	16-24	Female	Egypt	True	5	0-3	
1	16-24	Female	Egypt	True	4	0-3	
2	16-24	Male	Egypt	True	5	NaN	
3	16-24	Female	Egypt	False	1	NaN	
4	16-24	Female	Egypt	True	4	NaN	
5	16-24	Female	Egypt	True	4	NaN	

Data Visualization

Before coming up with any correlations, outliers need to be checked and dropped if found.

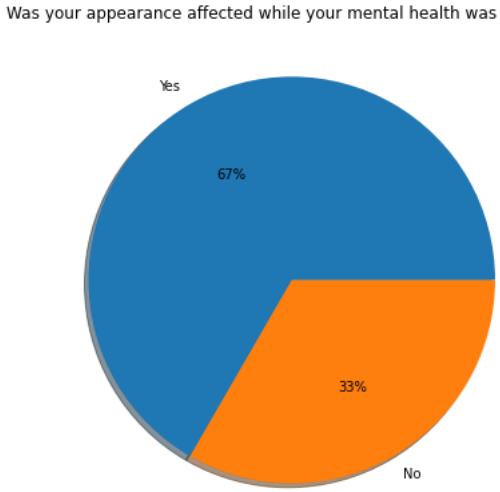
```
plt.figure(figsize=(15,7))
df["age"].value_counts().plot(kind = 'bar')
plt.show()
```



Age Range 45+ is an outlier and will be removed

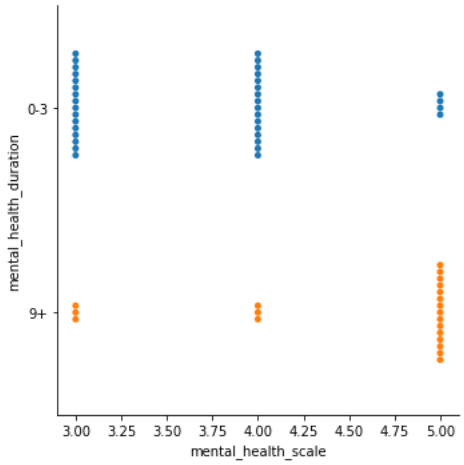
```
df.drop(df[df["age"] == "45+"].index, inplace=True)
plt.figure(figsize=(15,7))
df["age"].value_counts().plot(kind = 'bar')
plt.show()
```


```
fig = plt.figure(figsize=(10,7))
plt.title("Was your appearance affected while your mental health was affected?")
plt.pie(df["appearance_affected"].value_counts(), labels = ["Yes", "No"], shadow = True,autopct='%1.0f%%')
plt.show()
```




```
sns.catplot(data=df, x="mental_health_scale", y="mental_health_duration", kind="swarm")
```

```
<seaborn.axisgrid.FacetGrid at 0x7f93e0597d30>
```



According to the previous graph, it is shown that the individuals' rating on the 1-5 scale matched with their mental health issues' durations. As the scale goes up, duration increases.

```
sns.catplot(data=df, x="mental_health_duration", y="appearance_changed_count", hue="gender", kind="swarm")
```

```
<seaborn.axisgrid.FacetGrid at 0x7f93e2de0ac0>
```