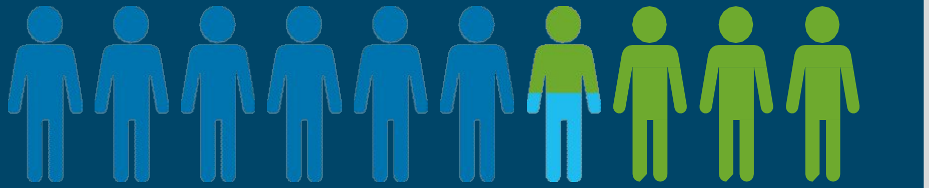
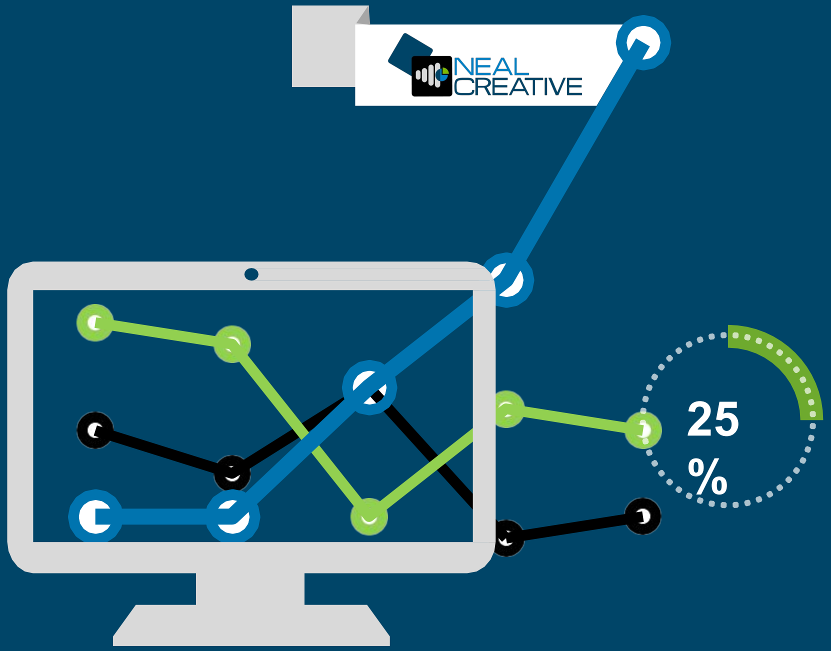
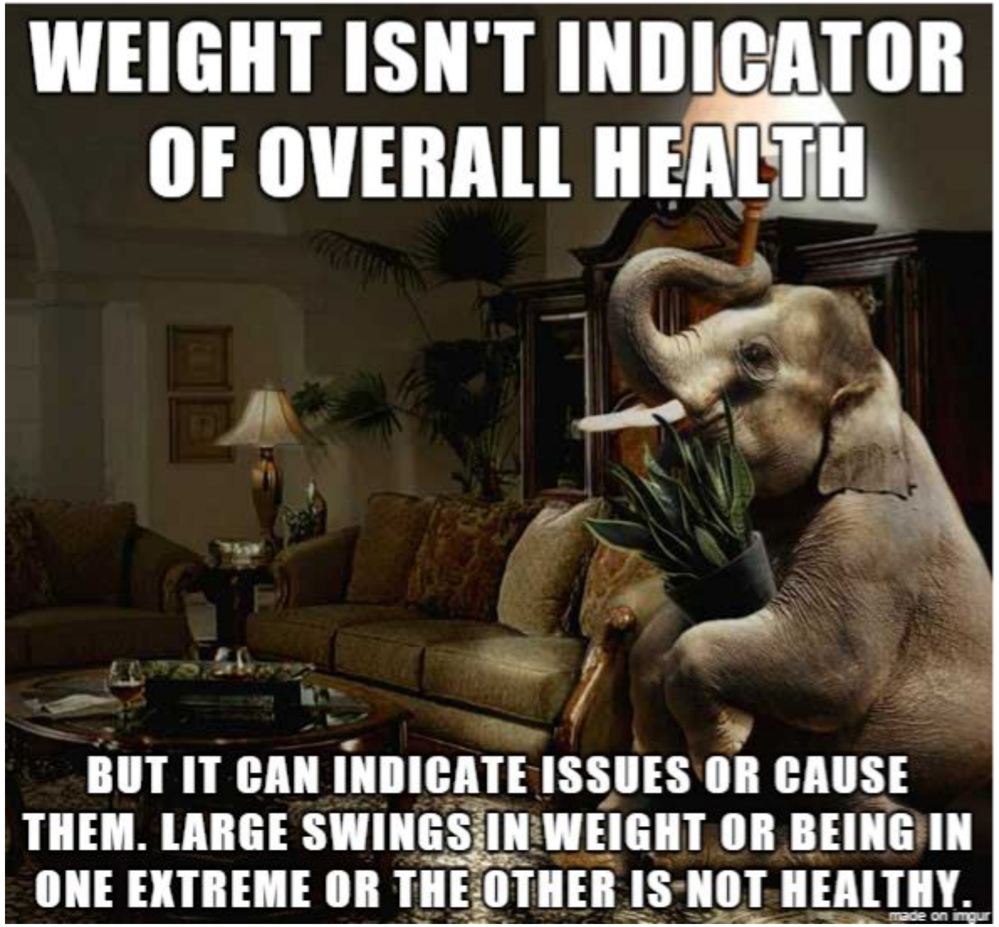
Prediction of Health Status Based on BMI

Remember to check your BMI...





HYPOTHESIS


WEIGHT ISN'T INDICATOR OF OVERALL HEALTH
BUT IT CAN INDICATE ISSUES OR CAUSE THEM. LARGE SWINGS IN WEIGHT OR BEING IN ONE EXTREME OR THE OTHER IS NOT HEALTHY.
made on imgur

Prediction of health status based on BMI

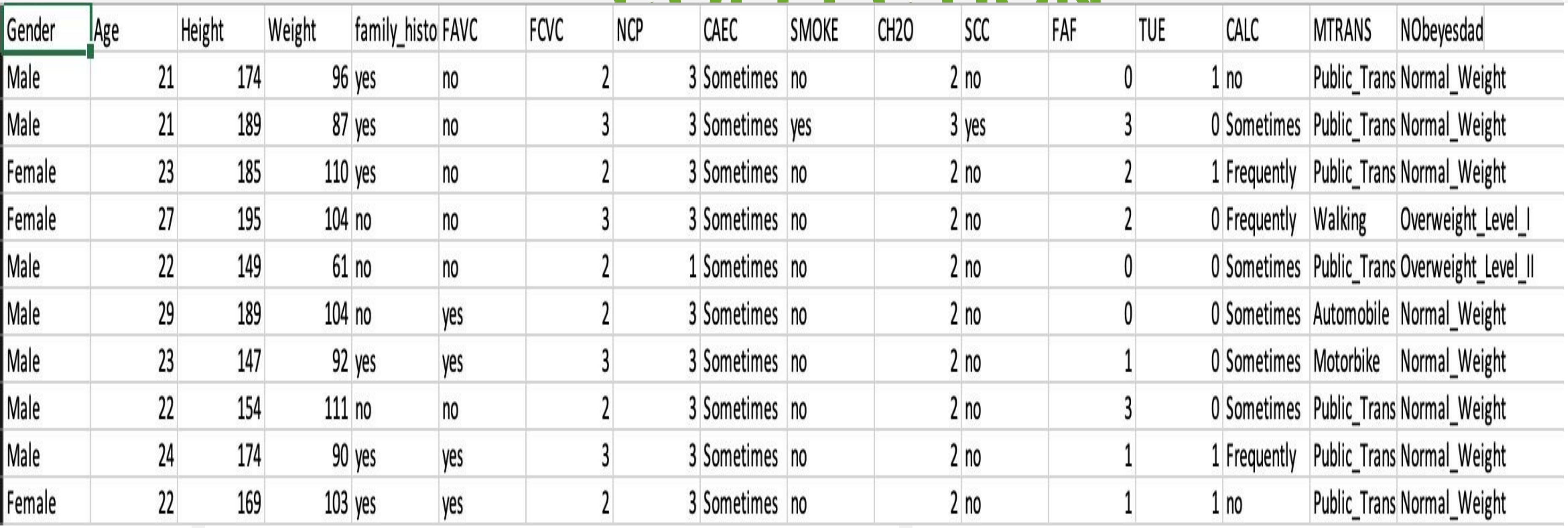
BMI prediction is constantly mocked by the media and the general public. "Is B.M.I. a

Scam?" is a question that most people and the media ask.

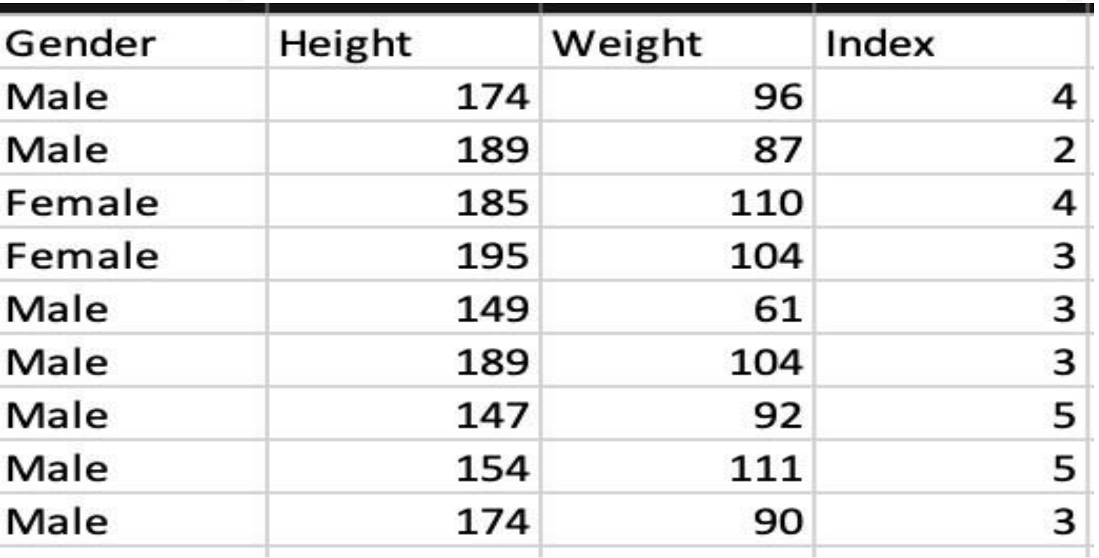
I would like to prove my hypothesis and demonstrate how beneficial BMI checking and

keeping track of your weight in line with your height is for a human being to live a healthy life. Using Kaggle datasets and machine learning models of random forest/linear regression to train existing datasets and do predictions for the new dataset.

DATA COLLECTION: RAW



Gender	Age	Height	Weight	family_histo	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
Male	21	174	96	yes	no	2	3	Sometimes	no	2	no	0	1	no	Public_Trans	Normal_Weight
Male	21	189	87	yes	no	3	3	Sometimes	yes	3	yes	3	0	Sometimes	Public_Trans	Normal_Weight
Female	23	185	110	yes	no	2	3	Sometimes	no	2	no	2	1	Frequently	Public_Trans	Normal_Weight
Female	27	195	104	no	no	3	3	Sometimes	no	2	no	2	0	Frequently	Walking	Overweight_Level_I
Male	22	149	61	no	no	2	1	Sometimes	no	2	no	0	0	Sometimes	Public_Trans	Overweight_Level_II
Male	29	189	104	no	yes	2	3	Sometimes	no	2	no	0	0	Sometimes	Automobile	Normal_Weight
Male	23	147	92	yes	yes	3	3	Sometimes	no	2	no	1	0	Sometimes	Motorbike	Normal_Weight
Male	22	154	111	no	no	2	3	Sometimes	no	2	no	3	0	Sometimes	Public_Trans	Normal_Weight
Male	24	174	90	yes	yes	3	3	Sometimes	no	2	no	1	1	Frequently	Public_Trans	Normal_Weight
Female	22	169	103	yes	yes	2	3	Sometimes	no	2	no	1	1	no	Public_Trans	Normal_Weight

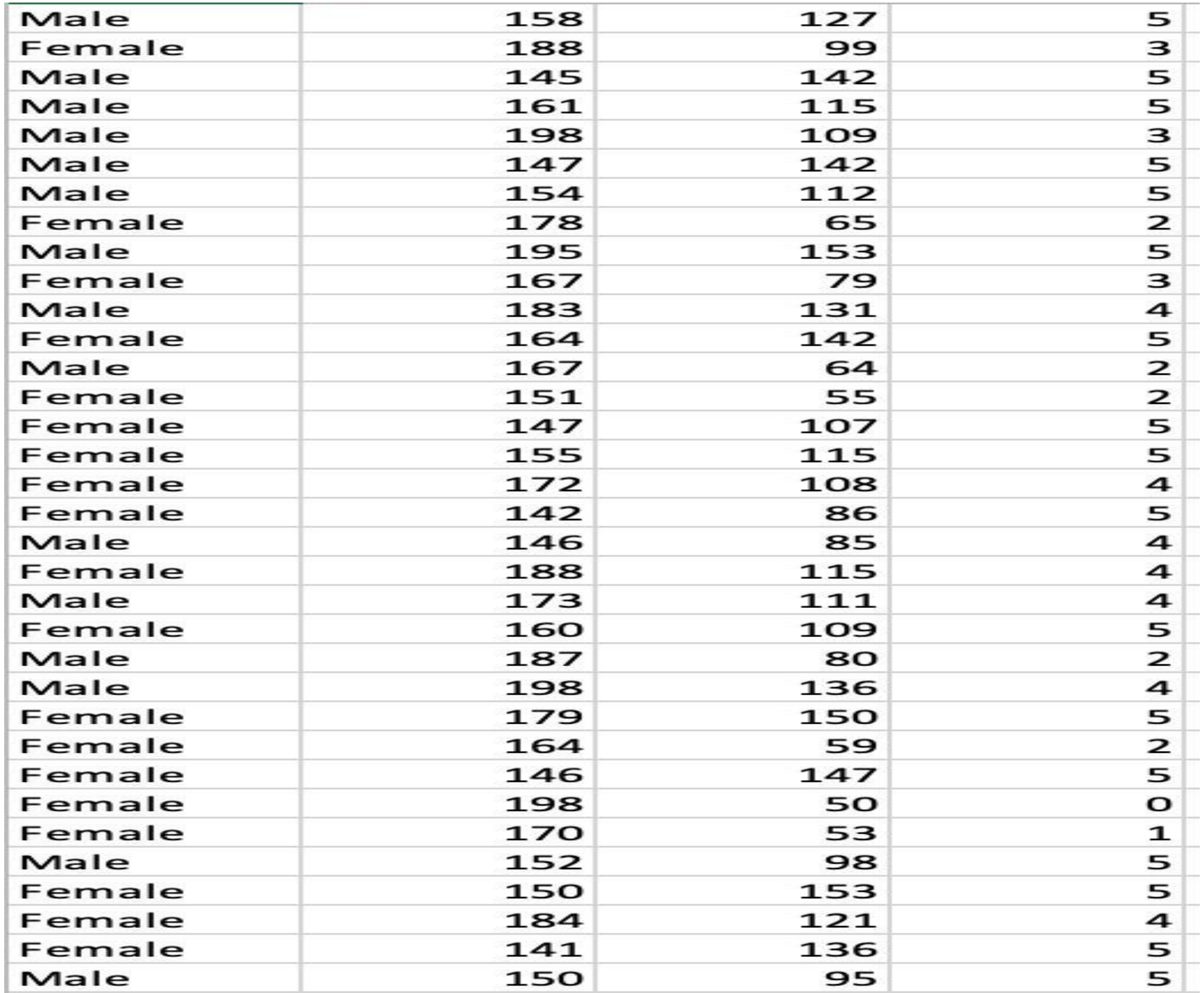


Gender	Height	Weight	Index
Male	174	96	4
Male	189	87	2
Female	185	110	4
Female	195	104	3
Male	149	61	3
Male	189	104	3
Male	147	92	5
Male	154	111	5
Male	174	90	3

Raw data

The original sources that the Kaggle dataset came from Pubmed.GOV, UC Machine Learning Repository. There is 19 attributes and 2111 rows in the original dataset; useful fields are person's gender, height, weight, and index.

DATA COLLECTION: CLEANED



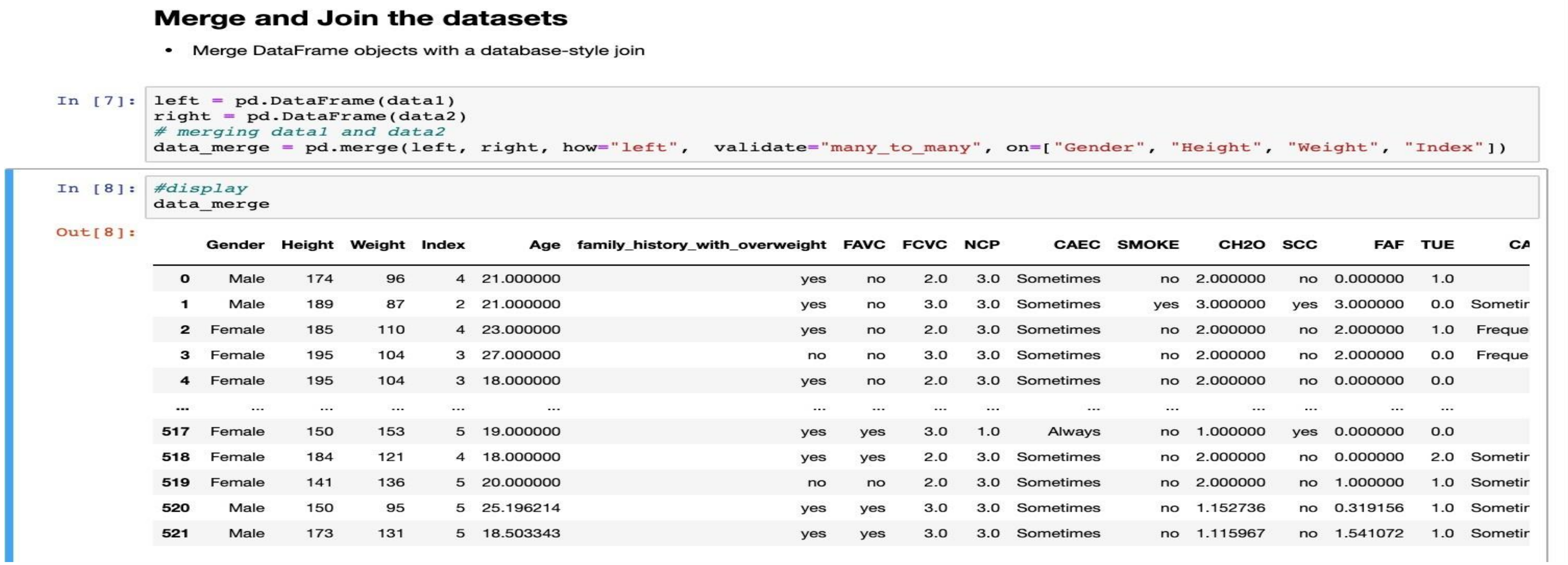
Male	158	127	5
Female	188	99	3
Male	145	142	5
Male	161	115	5
Male	198	109	3
Male	147	142	5
Male	154	112	5
Female	178	65	2
Male	195	153	5
Female	167	79	3
Male	183	131	4
Female	164	142	5
Male	167	64	2
Female	151	55	2
Female	147	107	5
Female	155	115	5
Female	172	108	4
Female	142	86	5
Male	146	85	4
Female	188	115	4
Male	173	111	4
Female	160	109	5
Male	187	80	2
Male	198	136	4
Female	179	150	5
Female	164	59	2
Female	146	147	5
Female	198	50	0
Female	170	53	1
Male	152	98	5
Female	150	153	5
Female	184	121	4
Female	141	136	5
Male	150	95	5

Dataset cleaned

Final Cleaned FOUR Columns: "Gender", "Height", "Weight", "Index"

Total = 500 rows; 4 columns

DATA FORMATTING: JUPTER NOTEBOOK



Merge and Join the datasets

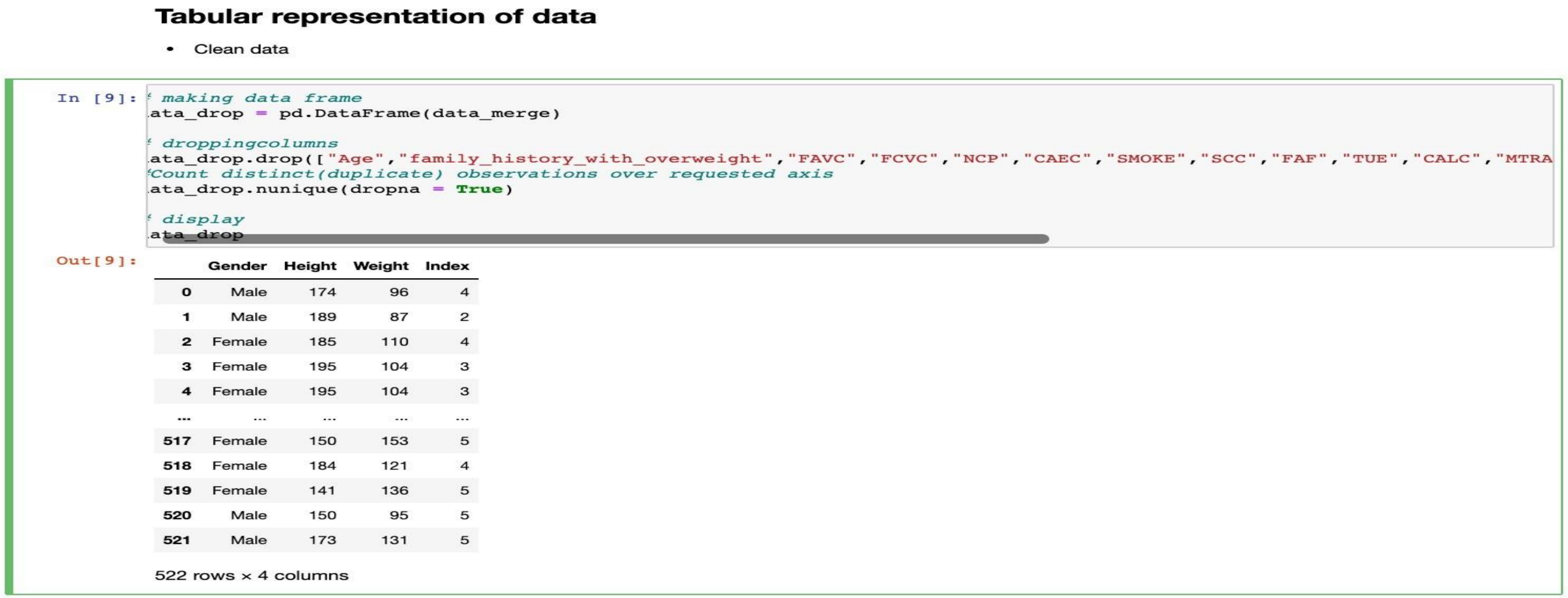
- Merge DataFrame objects with a database-style join

```
In [7]: left = pd.DataFrame(data1)
        right = pd.DataFrame(data2)
        # merging data1 and data2
        data_merge = pd.merge(left, right, how="left",  validate="many_to_many", on=["Gender", "Height", "Weight", "Index"])
```

```
In [8]: #display
        data_merge
```

Out[8]:

	Gender	Height	Weight	Index	Age	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CA
0	Male	174	96	4	21.000000	yes	no	2.0	3.0	Sometimes	no	2.000000	no	0.000000	1.0	
1	Male	189	87	2	21.000000	yes	no	3.0	3.0	Sometimes	yes	3.000000	yes	3.000000	0.0	Sometir
2	Female	185	110	4	23.000000	yes	no	2.0	3.0	Sometimes	no	2.000000	no	2.000000	1.0	Freque
3	Female	195	104	3	27.000000	no	no	3.0	3.0	Sometimes	no	2.000000	no	2.000000	0.0	Freque
4	Female	195	104	3	18.000000	yes	no	2.0	3.0	Sometimes	no	2.000000	no	0.000000	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
517	Female	150	153	5	19.000000	yes	yes	3.0	1.0	Always	no	1.000000	yes	0.000000	0.0	
518	Female	184	121	4	18.000000	yes	yes	2.0	3.0	Sometimes	no	2.000000	no	0.000000	2.0	Sometir
519	Female	141	136	5	20.000000	no	no	2.0	3.0	Sometimes	no	2.000000	no	1.000000	1.0	Sometir
520	Male	150	95	5	25.196214	yes	yes	3.0	3.0	Sometimes	no	1.152736	no	0.319156	1.0	Sometir
521	Male	173	131	5	18.503343	yes	yes	3.0	3.0	Sometimes	no	1.115967	no	1.541072	1.0	Sometir



Tabular representation of data

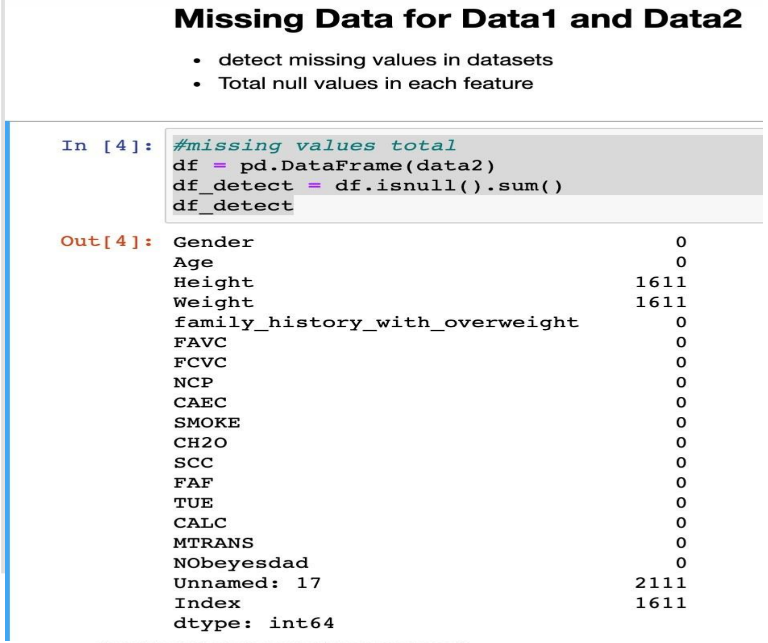
- Clean data

```
In [9]: ‘ making data frame
        ata_drop = pd.DataFrame(data_merge)

        ‘ droppingcolumns
        ata_drop.drop(["Age","family_history_with_overweight","FAVC","FCVC","NCP","CAEC","SMOKE","SCC","FAF","TUE","CALC","MTRA
        ‘Count distinct(duplicate) observations over requested axis
        ata_drop.nunique(dropna = True)

        ‘ display
        ata_drop
```

Out[9]:

	Gender	Height	Weight	Index
0	Male	174	96	4
1	Male	189	87	2
2	Female	185	110	4
3	Female	195	104	3
4	Female	195	104	3
...	...	...	...	...
517	Female	150	153	5
518	Female	184	121	4
519	Female	141	136	5
520	Male	150	95	5
521	Male	173	131	5

522 rows × 4 columns

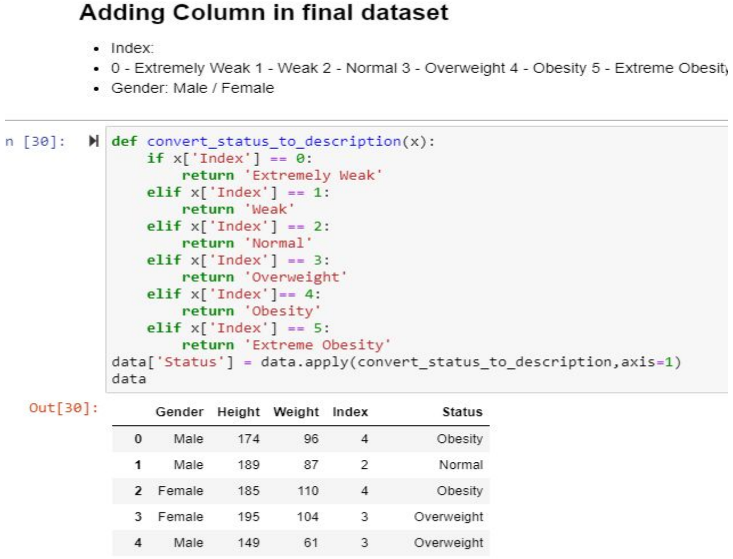


Missing Data for Data1 and Data2

- detect missing values in datasets
- Total null values in each feature

```
In [4]: #missing values total
        df = pd.DataFrame(data2)
        df_detect = df.isnull().sum()
        df_detect
```

```
Out[4]: Gender                                0
        Age                                   0
        Height                             1611
        Weight                             1611
        family_history_with_overweight        0
        FAVC                                  0
        FCVC                                  0
        NCP                                   0
        CAEC                                  0
        SMOKE                                 0
        CH2O                                  0
        SCC                                   0
        FAF                                   0
        TUE                                   0
        CALC                                  0
        MTRANS                                0
        NObeyesdad                            0
        Unnamed: 17                        2111
        Index                              1611
        dtype: int64
```



Adding Column in final dataset

- Index:
- 0 - Extremely Weak 1 - Weak 2 - Normal 3 - Overweight 4 - Obesity 5 - Extreme Obesit
- Gender: Male / Female

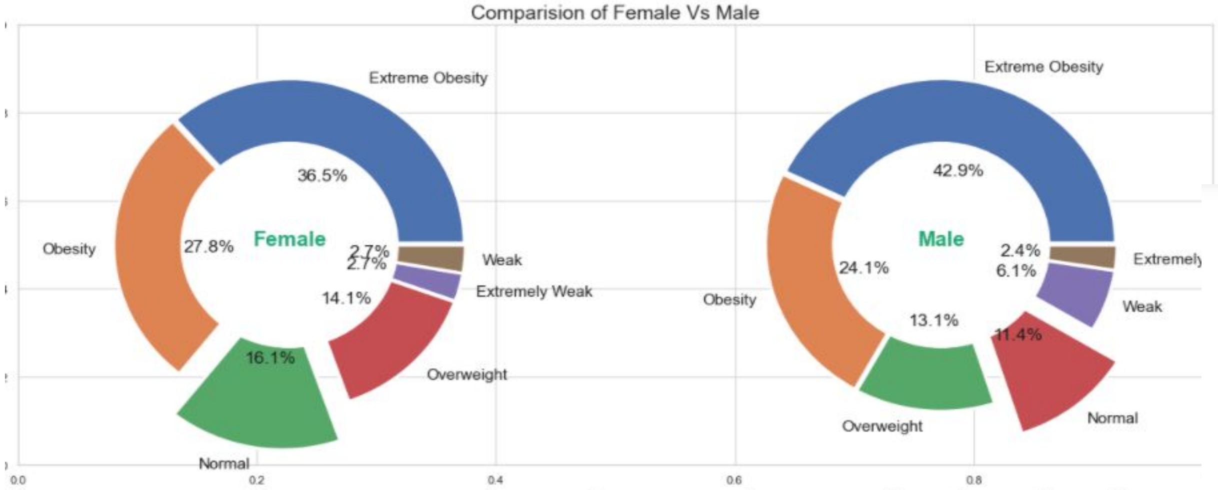
```
n [30]: def convert_status_to_description(x):
            if x['Index'] == 0:
                return 'Extremely Weak'
            elif x['Index'] == 1:
                return 'Weak'
            elif x['Index'] == 2:
                return 'Normal'
            elif x['Index'] == 3:
                return 'Overweight'
            elif x['Index']== 4:
                return 'Obesity'
            elif x['Index'] == 5:
                return 'Extreme Obesity'
        data['Status'] = data.apply(convert_status_to_description,axis=1)
        data
```

Out[30]:

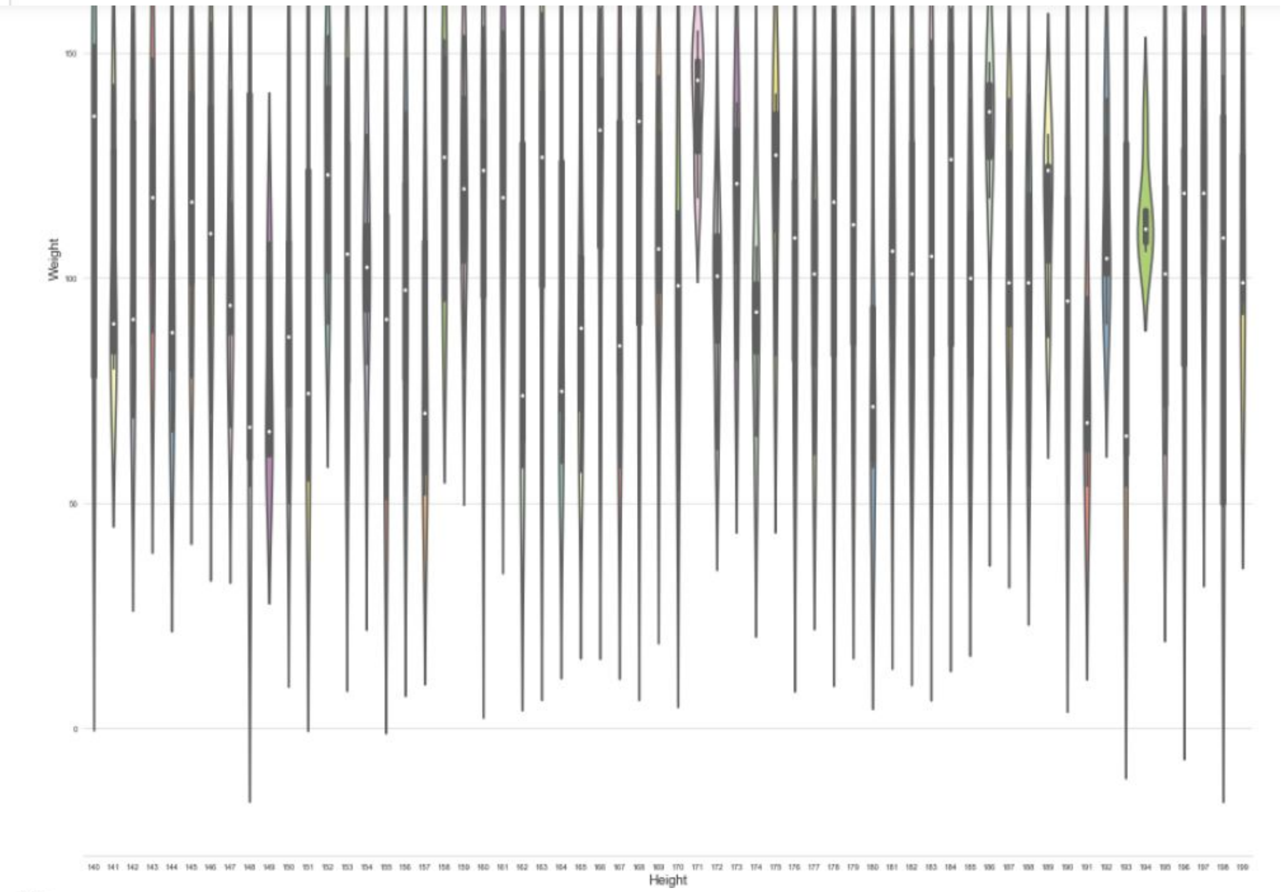
	Gender	Height	Weight	Index	Status
0	Male	174	96	4	Obesity
1	Male	189	87	2	Normal
2	Female	185	110	4	Obesity
3	Female	195	104	3	Overweight
4	Male	149	61	3	Overweight

DATA FORMATTING: JUPTER NOTEBOOK

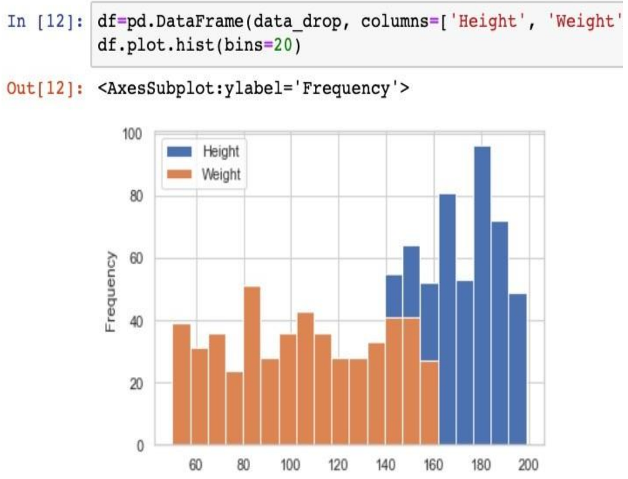
Pie-Plot Comparision of Female Vs Male



Violin plot visualization

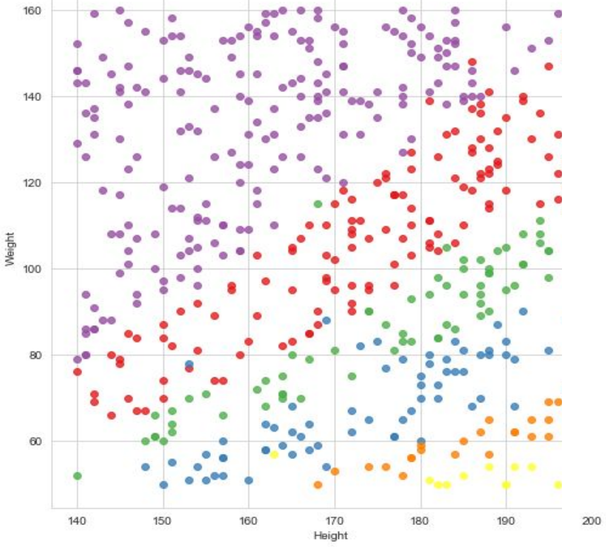


histogram (bar chart) visualization



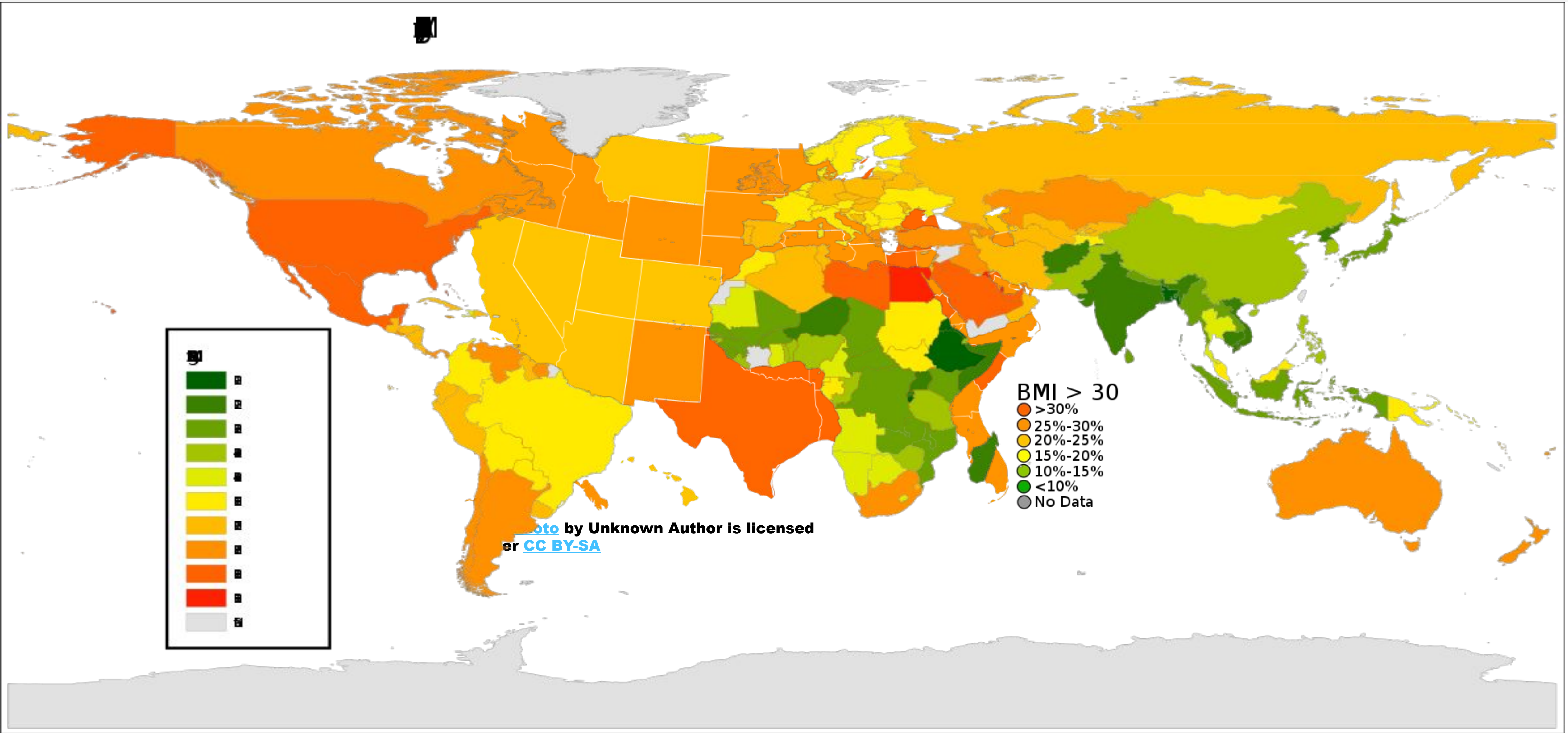
```
In [12]: df=pd.DataFrame(data_drop, columns=['Height', 'Weight'
         df.plot.hist(bins=20)

Out[12]: <AxesSubplot:ylabel='Frequency'>
```

Scatterplot matrix visualization



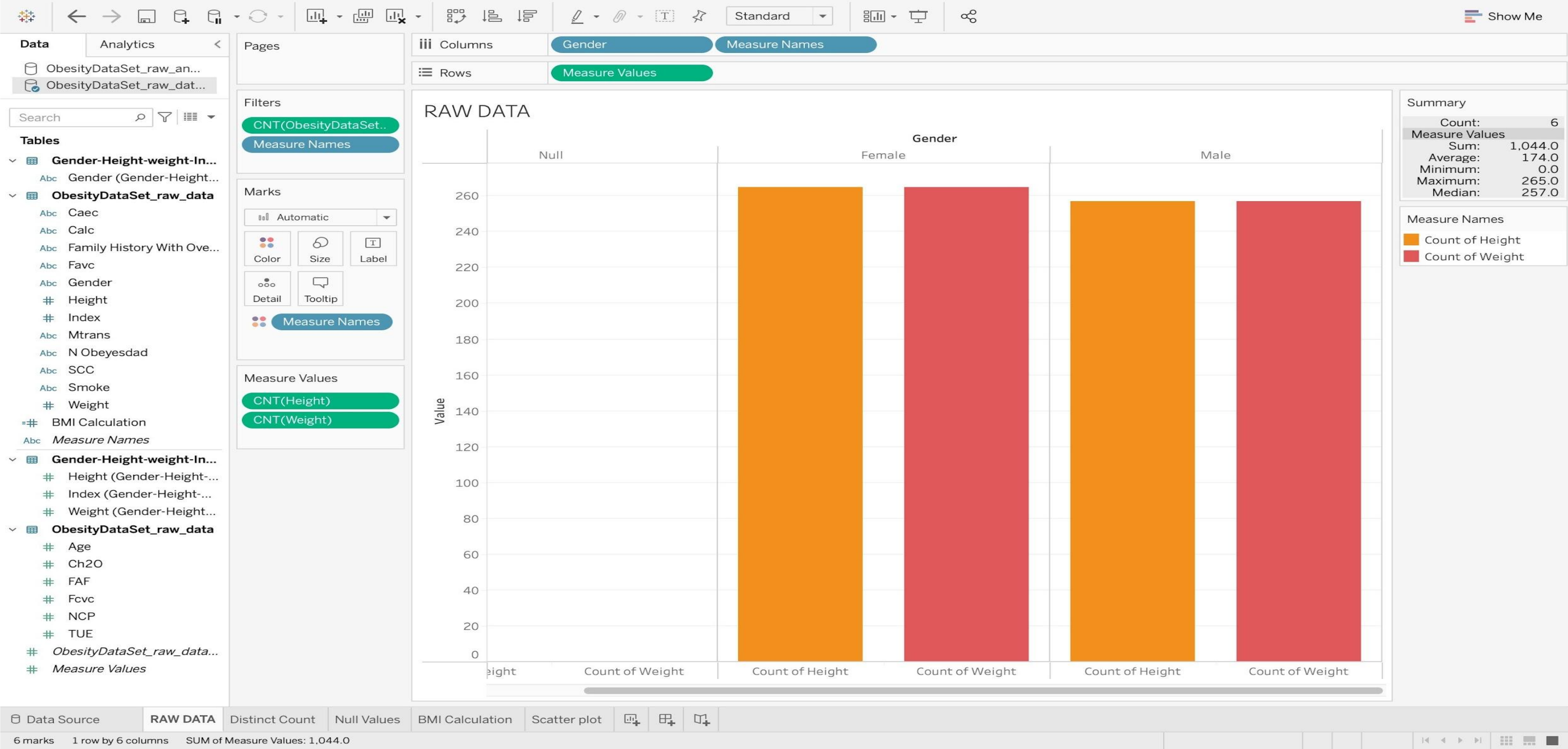
DATA VISUALITION

WORLD BMI VISUAL



RAW DATA SUMMARY

Male Vs Female



Columns: Gender, Measure Names

Rows: Measure Values

Filters: CNT(ObesityDataSet.., Measure Names

Marks: Automatic; Color: Measure Names

Measure Values: CNT(Height), CNT(Weight)

RAW DATA

Summary

Count:	6
Measure Values	
Sum:	1,044.0
Average:	174.0
Minimum:	0.0
Maximum:	265.0
Median:	257.0

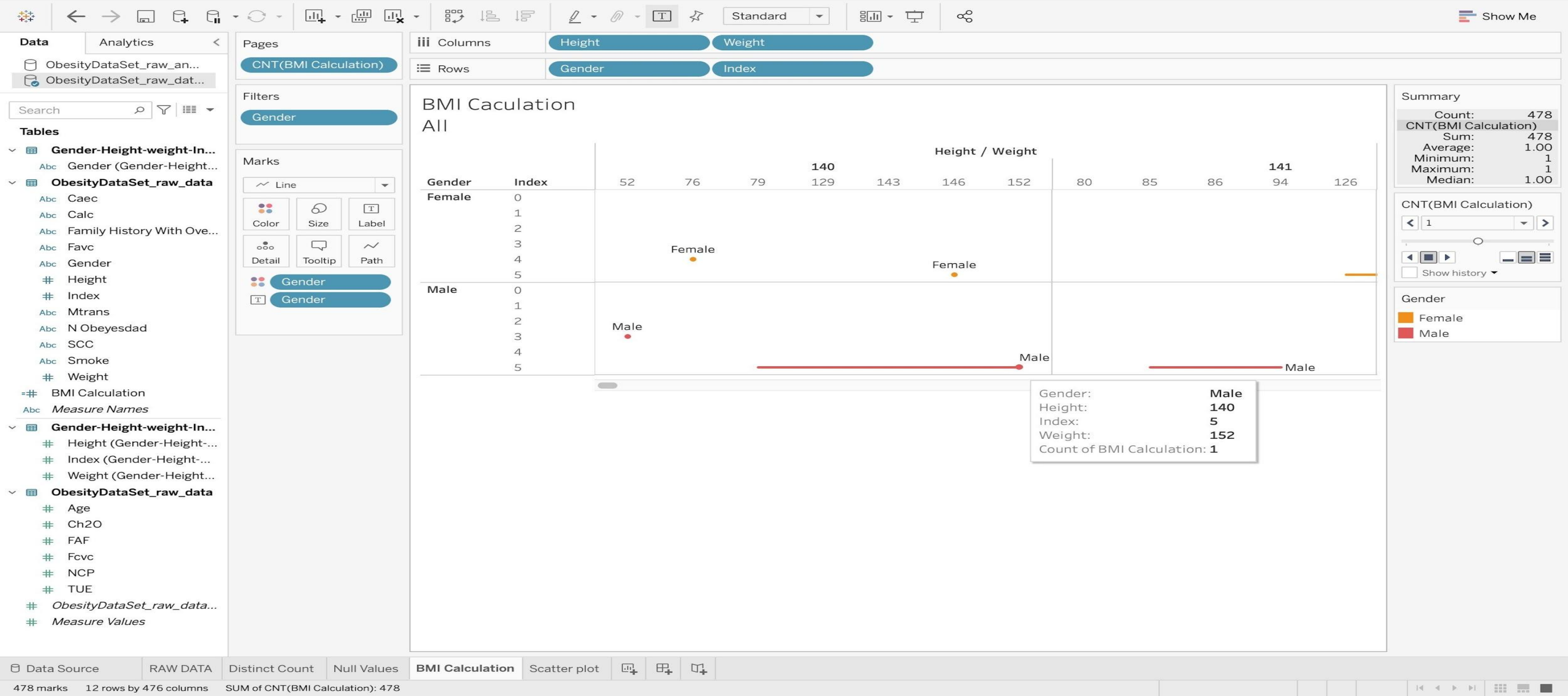
Measure Names

- 🟧 Count of Height
- 🟥 Count of Weight

6 marks 1 row by 6 columns SUM of Measure Values: 1,044.0

BMI CALCULATION

BMI = weight(kg) / height(cm)* height(cm)
Index= scaling BMI
Index: 0 - Extremely Weak 1 - Weak 2 - Normal 3 - Overweight 4 - Obesity 5 - Extreme Obesity
Gender: Male / Female



Pages: CNT(BMI Calculation)

Filters: Gender

Marks: Line
Gender
Gender

Columns: Height, Weight
Rows: Gender, Index

BMI Caculation
All

Summary

Count:	478
CNT(BMI Calculation)	
Sum:	478
Average:	1.00
Minimum:	1
Maximum:	1
Median:	1.00

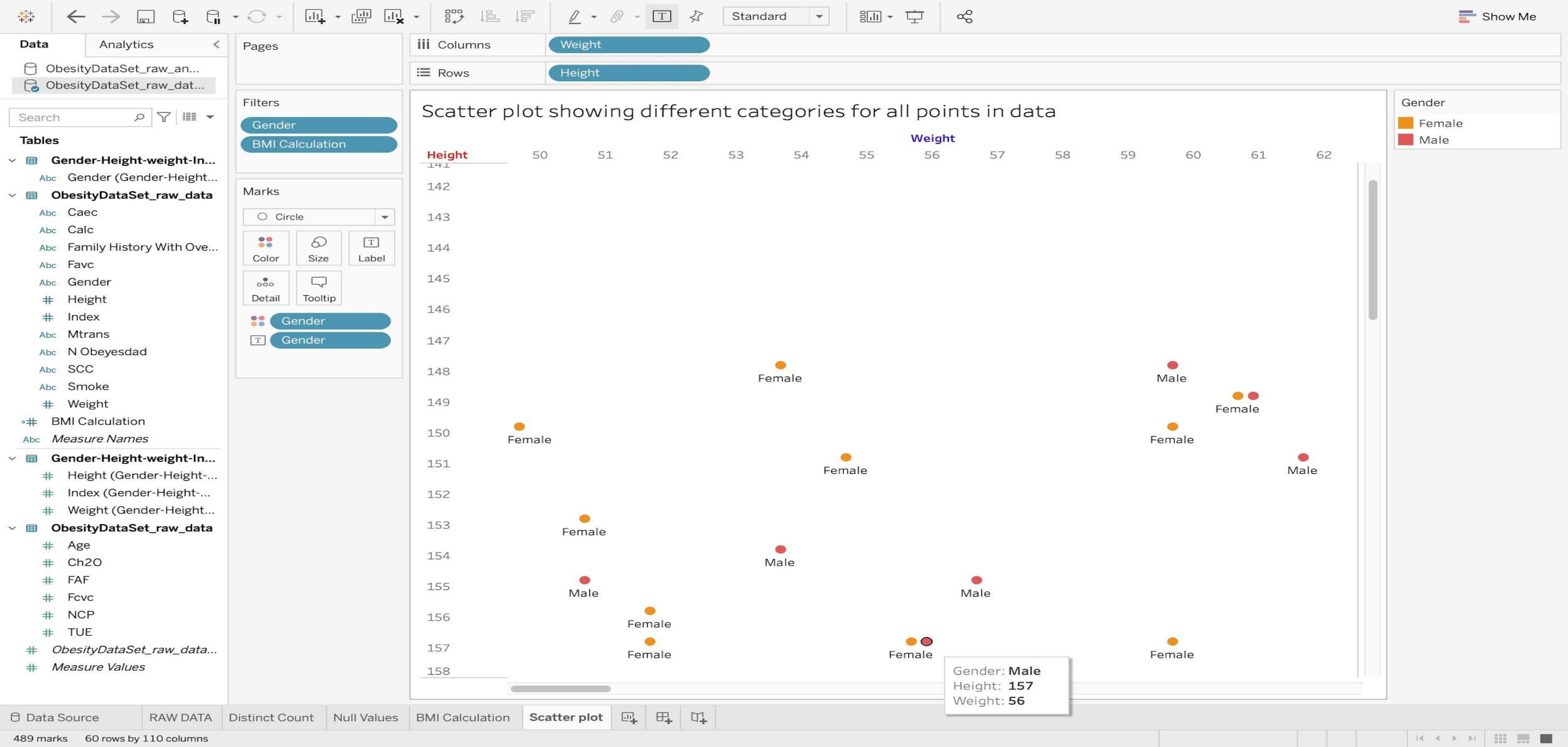
CNT(BMI Calculation): 1
□ Show history

Gender
🟧 Female
🟥 Male

Gender:	Male
Height:	140
Index:	5
Weight:	152
Count of BMI Calculation:	1

478 marks 12 rows by 476 columns SUM of CNT(BMI Calculation): 478

SCATTER PLOT: DIFFERENT CATEGORIES FOR ALL POINTS IN DATA



Data | Analytics

ObesityDataSet_raw_an...
ObesityDataSet_raw_dat...

Tables

- Gender-Height-weight-In...
  - Abc Gender (Gender-Height...
- ObesityDataSet_raw_data
  - Abc Caec
  - Abc Calc
  - Abc Family History With Ove...
  - Abc Favc
  - Abc Gender
  - # Height
  - # Index
  - Abc Mtrans
  - Abc N Obeyesdad
  - Abc SCC
  - Abc Smoke
  - # Weight
- BMI Calculation
- Abc *Measure Names*
- Gender-Height-weight-In...
  - # Height (Gender-Height-...
  - # Index (Gender-Height-...
  - # Weight (Gender-Height...
- ObesityDataSet_raw_data
  - # Age
  - # Ch2O
  - # FAF
  - # Fcvc
  - # NCP
  - # TUE
- # *ObesityDataSet_raw_data...*
- # *Measure Values*

Pages

Columns: Weight

Rows: Height

Filters: Gender, BMI Calculation

Marks: Circle

Color: Gender

Label: Gender

Scatter plot showing different categories for all points in data

Gender: Male
Height: 157
Weight: 56

Data Source | RAW DATA | Distinct Count | Null Values | BMI Calculation | Scatter plot

489 marks 60 rows by 110 columns

CONCLUSION

Concluding of Hypothesis

- Finally, based on the BMI hypothesis, you can predict your health state.
- Weight, according to the statistics, is a good determinant of overall health.
- To live a healthy life, a human being must check their BMI on a scale of 0-5 and keep track of their weight in relation to their height.
- In this research, bigdata analysis assists us in determining the appropriate BMI index scale for any gender.
- Exposure of data: I learned more about data preparation, such as merging, cleansing, and male/female classification.
- To demonstrate in a visual effect in order to gain a better understanding of the data.
- As a result, our hypothesis has been validated in this instance.

REFERENCES

Obesity Dataset Raw and Data Synthetic:
https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+ physical+condition+

https://pubmed.ncbi.nlm.nih.gov/12942320/

https://pubmed.ncbi.nlm.nih.gov/26036702/

https://pubmed.ncbi.nlm.nih.gov/23619660/

DOI: 10.1007/s00431-003-1292-x

DOI: 10.1093/pubmed/fdv067