Metabolic Syndrome

Project Week 3

Elisa Cerdá Doñate, Carmen Matos

What is Metabolic Syndrome?

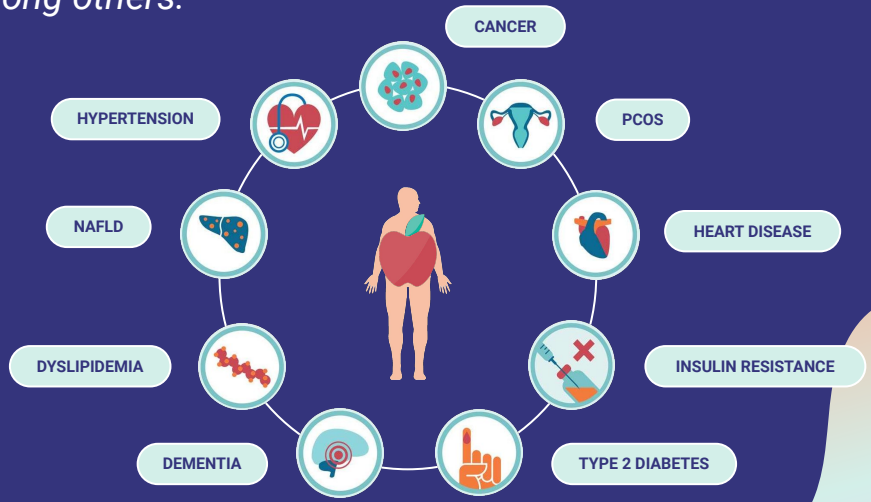
Metabolic syndrome refers to the presence of a cluster of risk factors specific for cardiovascular disease.

*Metabolic syndrome greatly **raises the risk of developing serious conditions**, such as diabetes or heart disease, among others.*

HDL: High density cholesterol
NAFLD: Non-alcoholic fatty liver disease
PCOS: Polycystic ovary syndrome

Diagnostic criteria (at least 3):

- Abdominal obesity
- High blood pressure
- Impaired fasting blood glucose
- High triglyceride levels
- Low HDL cholesterol



Sources: [1] Metabolic Syndrome. Johns Hopkins Medicine. *Accessed: 02 November 2023*. [2] Metabolic Syndrome: Are You At Risk? InBodyUSA. *Accessed: 02 November 2023*.

Hypotheses

- Metabolic Syndrome is more prominent in individuals above 50 years
- Metabolic Syndrome is more prominent among separated or divorced people
- Metabolic syndrome is more common among lower income groups
- Metabolic Syndrome is strongly correlated with waist circumferences and BMI values

Dataset and Data Cleaning

Our Dataset - Metabolic Syndrome (Kaggle)

- This dataset contains information on individuals with metabolic syndrome on a single table
- The data includes demographic, clinical, and laboratory measurements, as well as the presence or absence of metabolic syndrome

	seqn	Age	Sex	Marital	Income	Race	WaistCirc	BMI	Albuminuria	UrAlbCr	UricAcid	BloodGlucose	HDL	Triglycerides	MetabolicSyndrome
0	62161	22	Male	Single	8200.0	White	81.0	23.3	0	3.88	4.9	92	41	84	0
1	62164	44	Female	Married	4500.0	White	80.1	23.2	0	8.55	4.5	82	28	56	0
2	62169	21	Male	Single	800.0	Asian	69.6	20.1	0	5.07	5.4	107	43	78	0
3	62172	43	Female	Single	2000.0	Black	120.4	33.3	0	5.22	5.0	104	73	141	0
4	62177	51	Male	Married	NaN	Asian	81.1	20.1	0	8.13	5.0	95	43	126	0

Sources: [1] Metabolic Syndrome. Dataset. Kaggle. *https://www.kaggle.com/datasets/antimoni/metabolic-syndrome*

Data Cleaning

ORIGINAL DATAFRAME

Shape: (2401, 15)

Check for Null Values

seqn	0
Age	0
Sex	0
Marital	208
Income	117
Race	0
WaistCirc	85
BMI	26
Albuminuria	0
UrAlbCr	0
UricAcid	0
BloodGlucose	0
HDL	0
Triglycerides	0
MetabolicSyndrome	0

dtype: int64

CLEANING STEPS

1. Marital null value drop
2. Income null value drop
3. Mean for null BMI values
4. Mean for null waist values
5. 3σ BMI outlier drop

CURATED DATAFRAME

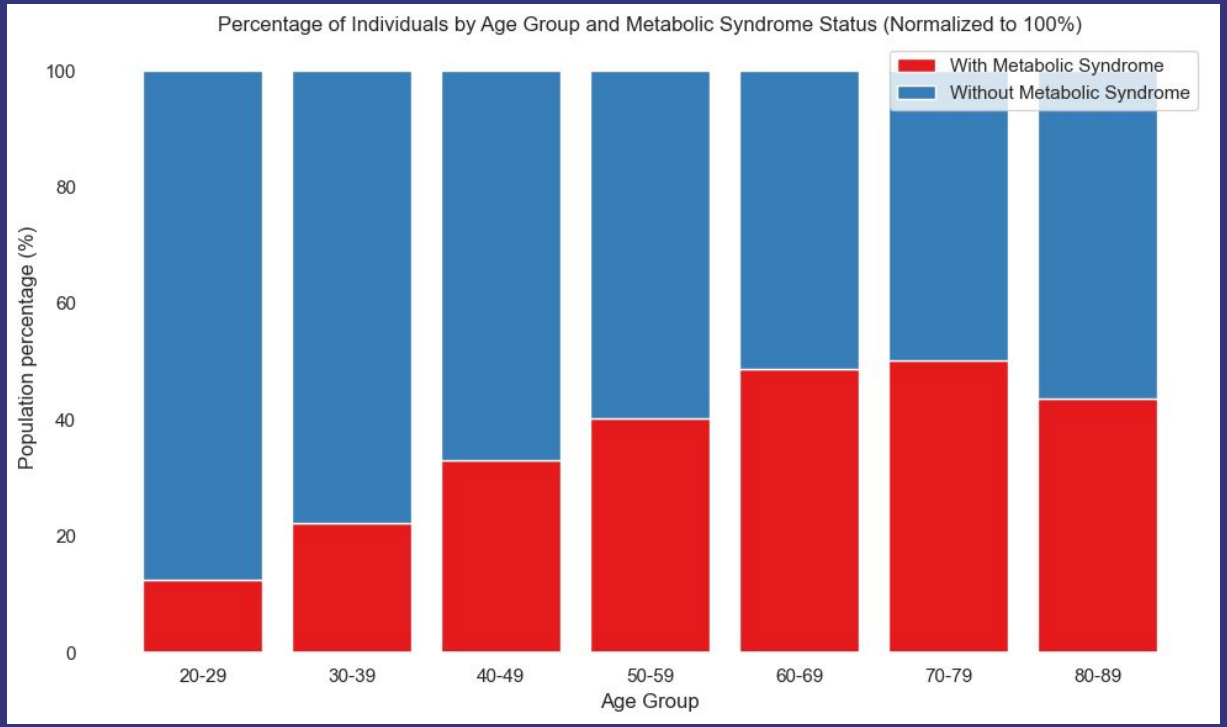
Shape: (2090, 15)

Check for Null Values

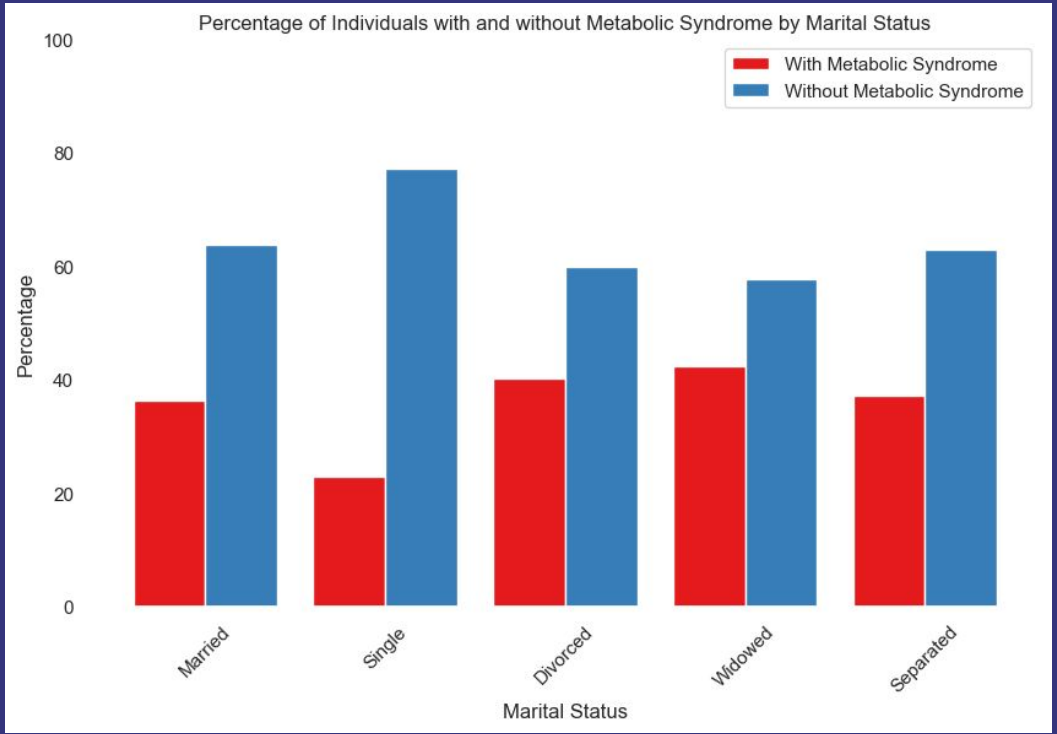
seqn	0
Age	0
Sex	0
Marital	0
Income	0
Race	0
WaistCirc	0
BMI	0
Albuminuria	0
UrAlbCr	0
UricAcid	0
BloodGlucose	0
HDL	0
Triglycerides	0
MetabolicSyndrome	0

dtype: int64

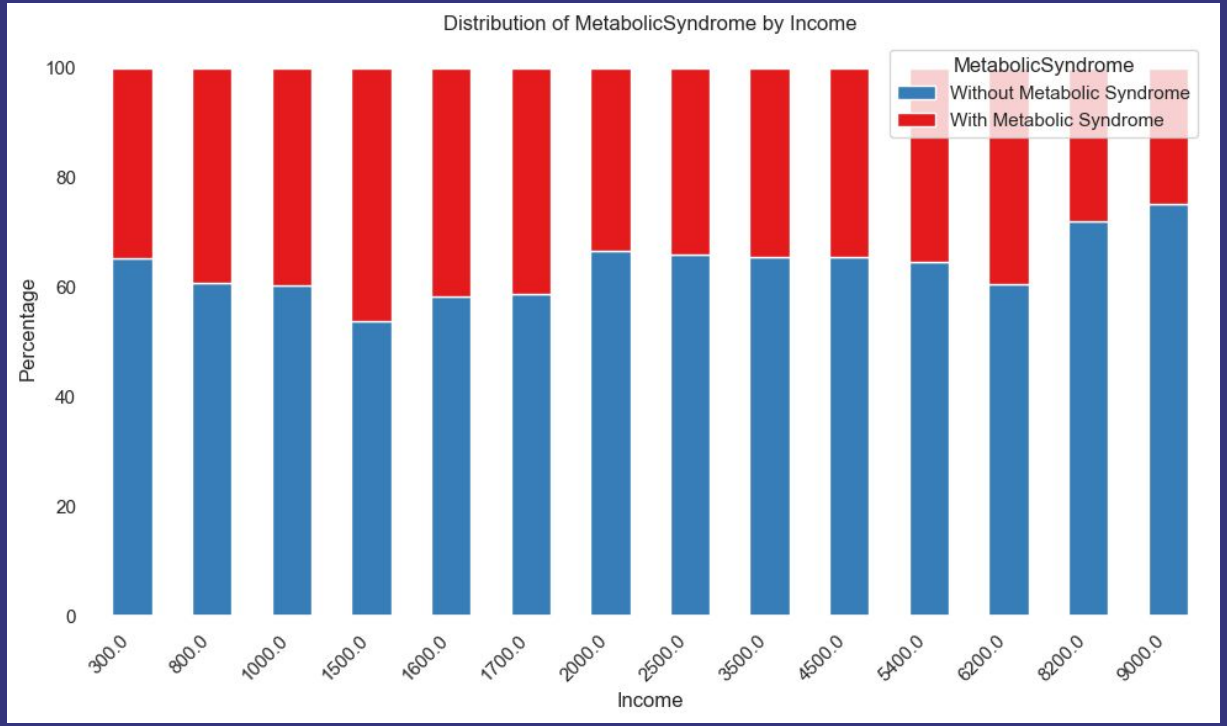
H1: Metabolic Syndrome is more prominent in individuals >50 years



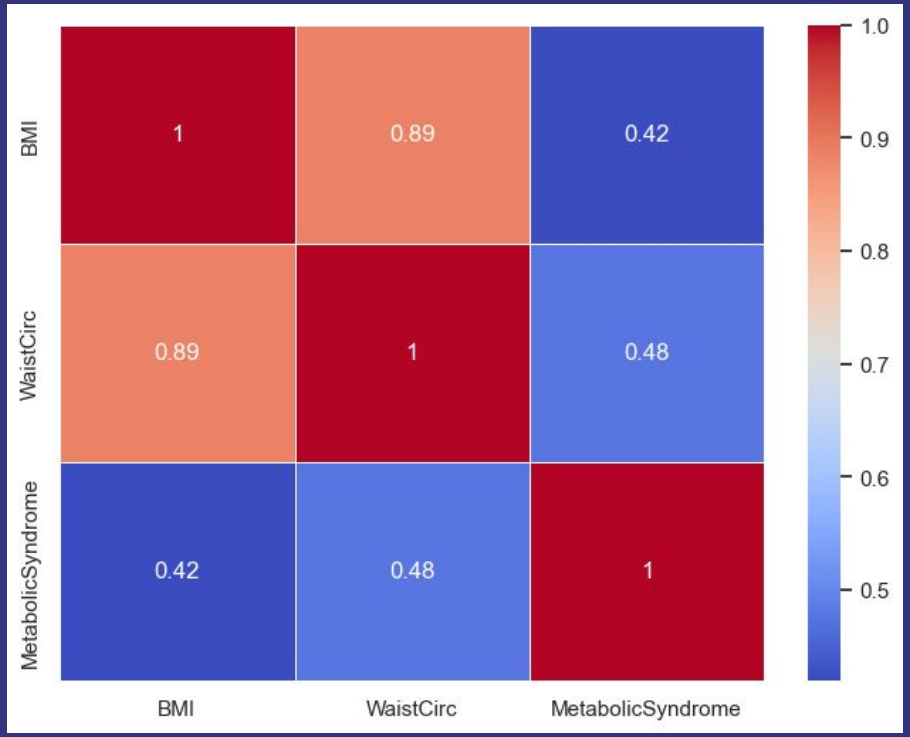
H2: Metabolic Syndrome is more prominent in separated or divorced people



H3: Metabolic Syndrome is more prominent among lower income groups



H4: Metabolic Syndrome is strongly correlated waist circumferences and BMI values



Challenges Faced

- **Data Cleaning** (Dropping and Filling Null Values and Outliers)
  - Had to consider how much values we could drop and how much it would affect our data
- **Data Normalization and Binning**
  - Took more time than expected on data normalization for visualization purposes
- **MySQL Use**
  - Wondered why using SQL instead of Python

Conclusion

- Our exploratory analysis showed a correlation between **presence of metabolic syndrome** and **BMI** and **waist circumference**
- Metabolic syndrome was found to become more prevalent with **increasing age**
- Health departments can make use of this type of data to determine **risk groups** and **develop prevention** and **intervention strategies** tailored to each risk group

Metabolic Syndrome

Project Week 3

Elisa Cerdá Doñate, Carmen Matos