DATE
20/4/2024

INSIGHTS INTO DRUG ABUSE
“EXPLORING BEHAVIORAL INDICATORS OF DRUG ABUSE”
DESCRIPTIVE DATA ANALYTICS APPROACH.

Data Analytics - Midterm - 2024

HTU
Leen Shareef Saleh
22110090

Course Tutor: Dr. Bassam Al-kasasbeh

DATA ANALYTIC ACTIVITIES AND TOOL

Data Visualization

Data Visualization helps making valuable decisions, recognizing trends, patterns and getting better understanding of the data, using several tools such as: line chart, bar chart, histogram, Boxplot.

Data Transformation

Data Transformation is the process of converting data from one form to another, for making it better for the analysis. this activity involves converting data types, normalizing numerical data and more. this activity is important for knowing what are the statistical measures that we can use them on the data, so we get actual analysis by choosing the correct methods and tools.

TOOL USED

Python

Python is a powerful language for data analysis, by having a group of libraires that could handle the manipulation of data easily.

- Pandas: loading data, understanding it with the structure of DataFrames or Series.
- Matplotlib: creating a wide range of visualizations
- LabelEncoder: it’s a class from the sklearn library machine that it was used for converting categorical data to numerical.
- seaborn: a library used for visualizations, statistical data visualization library, and It was used for the heatmap to get the correlation between the columns.

https://elearning.htu.edu.jo/pluginfile.php/98913/mod_resource/content/0/1.%20Theoretical%20foundation%20of%20data%20analytics.pdf

DATA ANAYLTICS METHODS

descriptive analytics

descriptive analytics answers the questions of what happened and why. by making summarize for the current and the historical data using statistical measures to get patterns, insights and trends. It is a method for preparation of a more advanced techniques.

Data modeling techniques used: data aggregation, summary statistics, and data visualization.

predictive analytics

This method tells what is likely to happen, by identifying possible outcomes and the probability that they will happen, based on the data that we already did descriptive analytics for.

- Anticipate most likely outcomes.
- forecast a sequence.
- make prediction based on an information that we have.
- identifying potential opportunities and risks ahead.

Note: there is always a degree of uncertainty associated with predictions.

Data modeling techniques: regression analysis, time series analysis, decision trees, and machine learning algorithms.

descriptive analytics

Descriptive analytics answers the question of what should we do? (what best to do?) determining the best course of action based on an outcome we want to achieve.

- make decisions based on historical data.
- to mitigate of possible risks.
- optimizing progress.

Data modeling techniques: more complex algorithms and optimization methods like linear programming, genetic algorithms, or simulation models.

FILE:///C:/USERS/LSHMJ/ONEDRIVE/DOCUMENTS/LEEN%20HTU/DATA%20ANALYTICS/DESCRIPTIVE%20ANALYTICS/2.%20DESCRIPTIVE%20ANALYTICS.PDF

THE USE OF THE THREE METHODS PROVIDING BY EXAMPLE

Drug Abuser

1) descriptive analytics: analyzing the historical data of the patients, (taken the drug before or not, Family History of Mental Illness, Age group, behavioral issues, individual's level of self-control, desire to use drugs and mental health), of the patients between the ages of 25 and 40, 40% had previously abused drugs like opioids, and 60% had previously abused alcohol. Second, 70% of patients who had a family history of mental illness also struggled with substance abuse. Our ability to better understand the patterns and dangers of substance abuse among patients will enable us to develop support programs that are more effective for them.

2) predictive analytics: looking at variables such as the individual's level of self-control, desire to use drugs and mental health. It predicts with a good degree of accuracy, about 80% of the time. For example, the program's findings indicate that approximately 85% of participants who show poor self-control and a strong desire for drugs are likely to return to drugs within the next six months. This facilitates early intervention by doctors and better treatment planning for the patient because he was predicted to have drugs by the percentage calculated.

3) perspective analytics: recommending a combination of therapies, such as talking therapy, medication, and mindfulness exercises if it determines that a patient has a high risk of recurrence. we will be taking action of monitoring the patient's development and modifying the course of treatment as necessary. Doctors may help patients in improving their general health and preventing the recurrence in drug use by implementing these actions.

https://www.mdpi.com/2076-3425/12/7/957

https://www.selecthub.com/business-intelligence/predictive-descriptive-prescriptive-analytics/

DESCRIPTIVE ANALYTICS TECHNIQUES

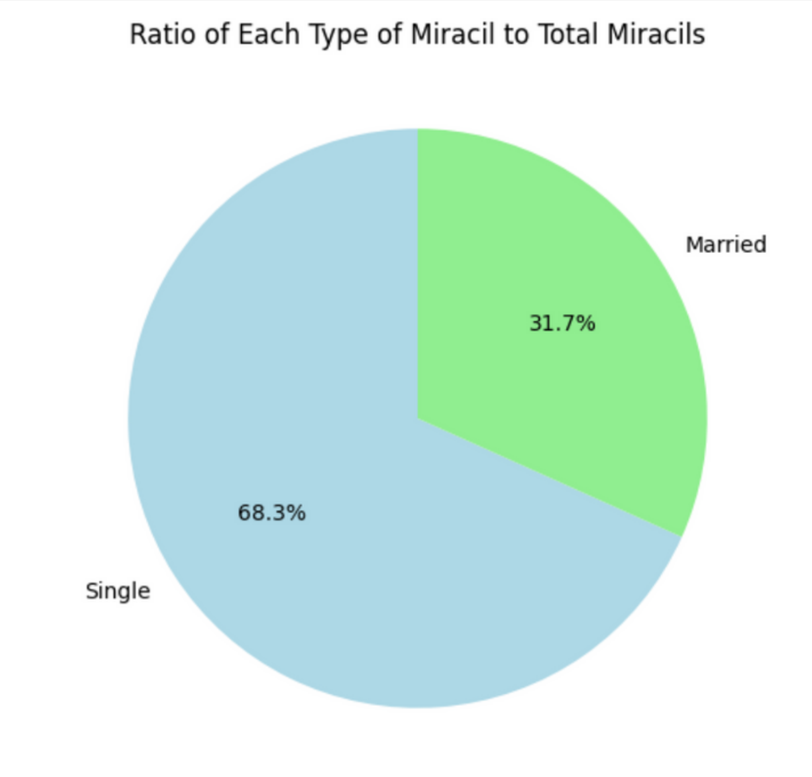
techniques:

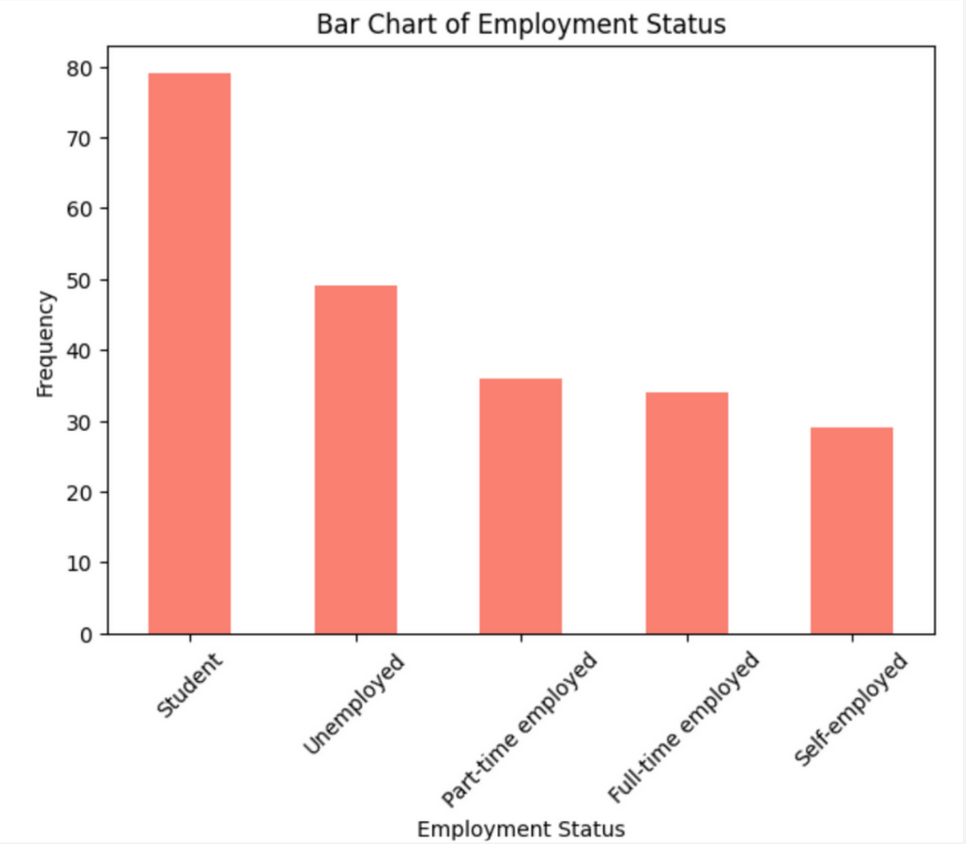
- measures of frequency: describes the amount of times a specific value appears in a dataset.
- Measures of Central Tendency: A single value that describes a dataset by finding its center
- Measures of Dispersion: finding out a dataset's variability or spread, showing the data points and their central tendency are separated. tendency.
- Measures of Position: statistical values that are used to describe where an individual data point or observation is located within a dataset.
- Contingency Table: Data is organized in a contingency table so that the relationship between two categorical variables can be shown. For every set of categories, it shows the frequencies or counts of observations.
- Outlier detections: Outlier detection is a process of identifying data points that are significantly different from the rest of the dataset.

BEFORE STARTING ANALYZING

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 227 entries, 0 to 226
Data columns (total 24 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Case_ID                         227 non-null    object
 1   Age                             227 non-null    int64
 2   Marital_Status                  227 non-null    object
 3   Employment_Status               227 non-null    object
 4   Income Level                    227 non-null    object
 5   Health Index                    227 non-null    int64
 6   Smoking                         227 non-null    object
 7   taken_drugs                     227 non-null    object
 8   absence_from_work               227 non-null    object
 9   sleeping_habits                 227 non-null    object
 10  headache                        227 non-null    object
 11  decreased_social_interaction    227 non-null    object
 12  increased_personal_expenses     227 non-null    object
 13  selling_house_content           227 non-null    object
 14  commit_stole                    227 non-null    object
 15  lying                           227 non-null    object
 16  car_crash                       227 non-null    object
 17  slow_reactions                  227 non-null    object
 18  weight_changed                  227 non-null    object
 19  distracted_and_absent_minded    227 non-null    object
 20  Weighted_Stress_Index           227 non-null    float64
 21  Risk_of_Legal_Issues            227 non-null    int64
 22  Risk_Factor_Score               227 non-null    float64
 23  Drug_or_not                     227 non-null    object
dtypes: float64(2), int64(3), object(19)
memory usage: 42.7+ KB
```

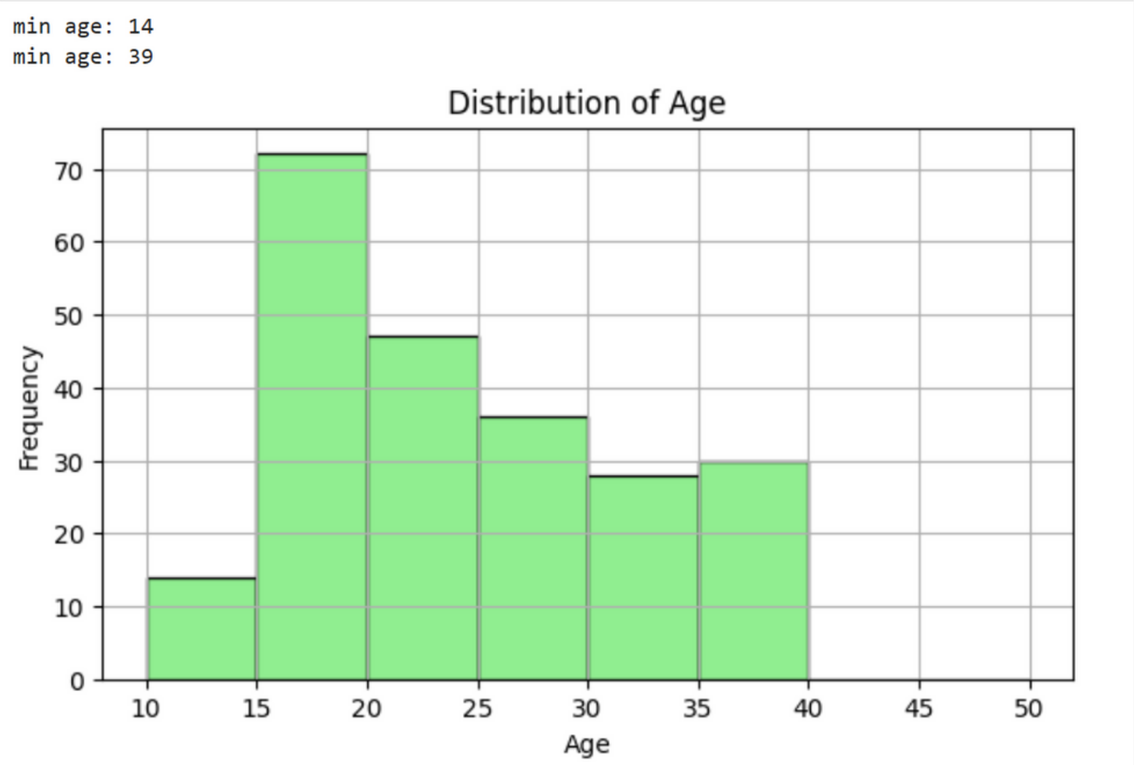
REGENERAL UNDERSTANDINGS

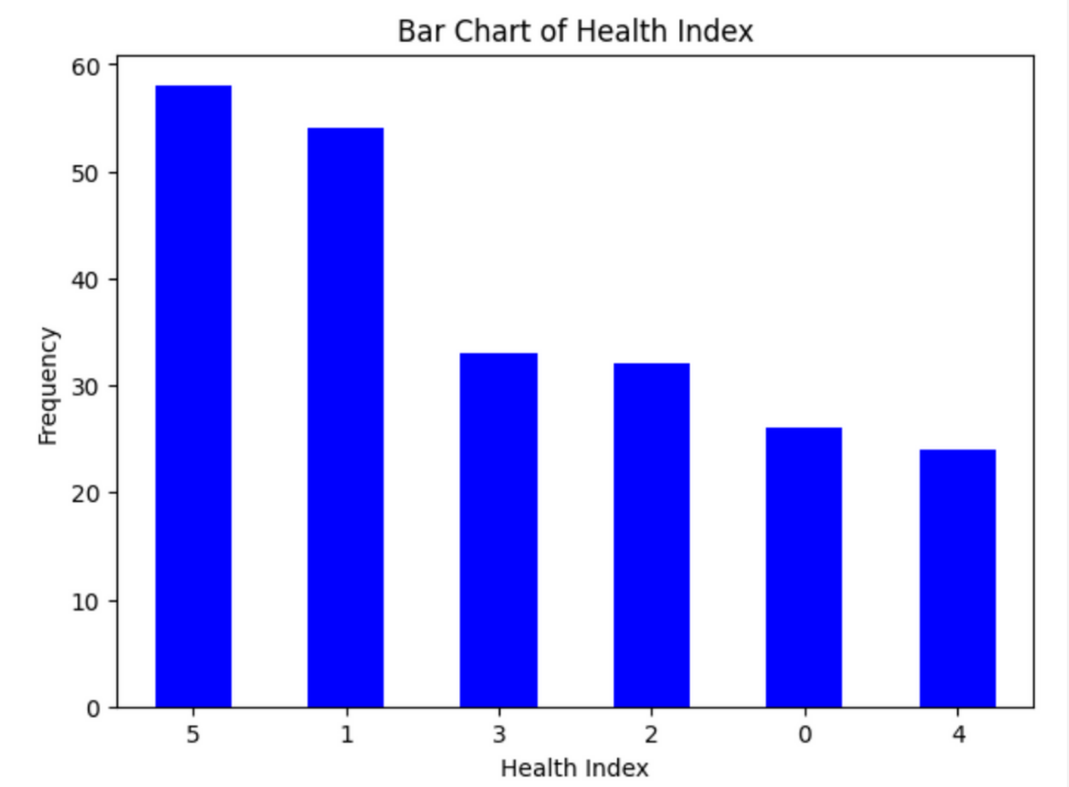




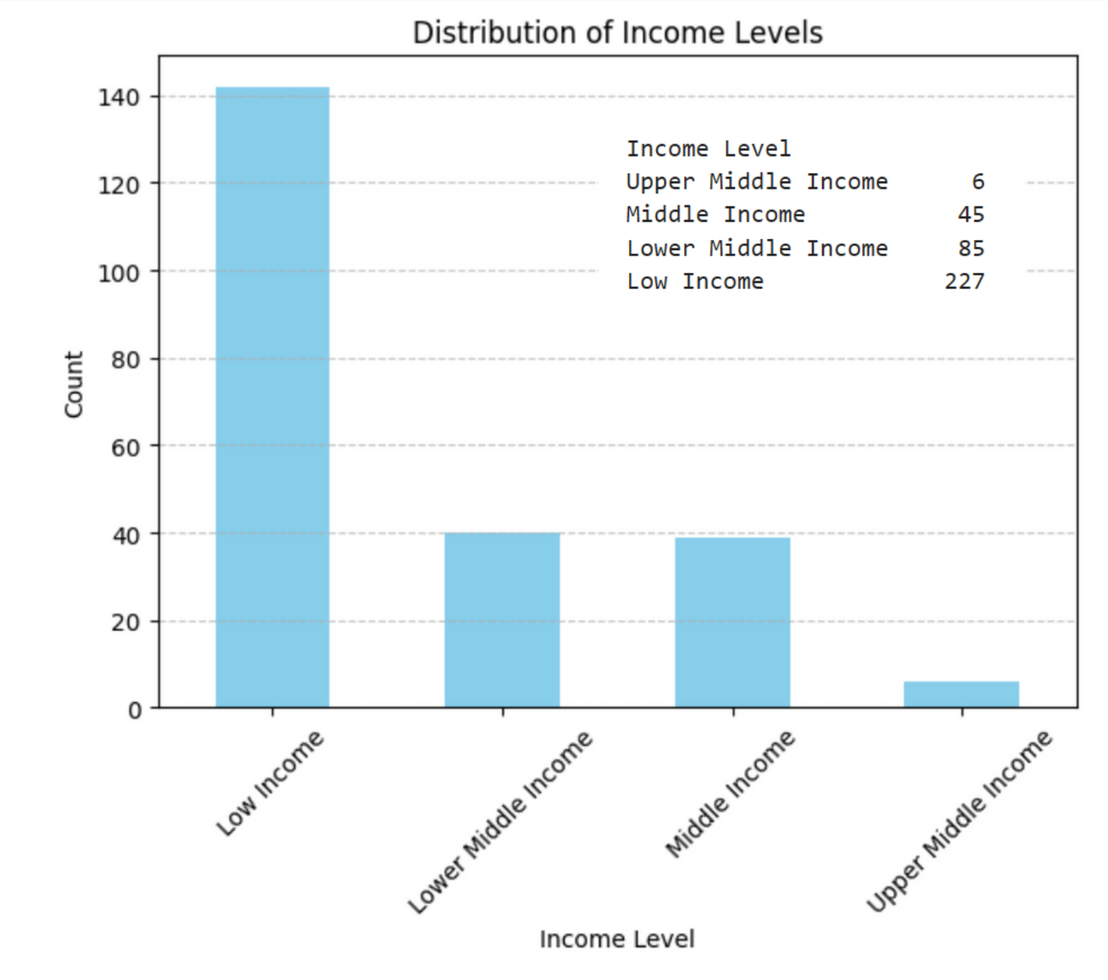
REGENERAL UNDERSTANDINGS

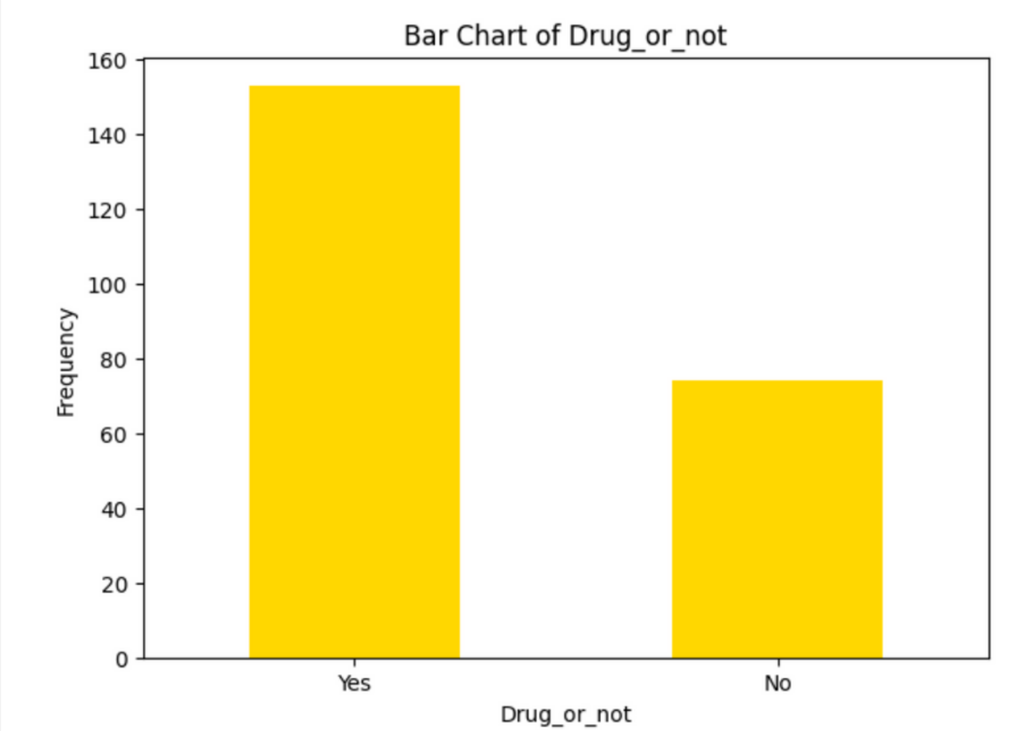
```
min age: 14
min age: 39
```





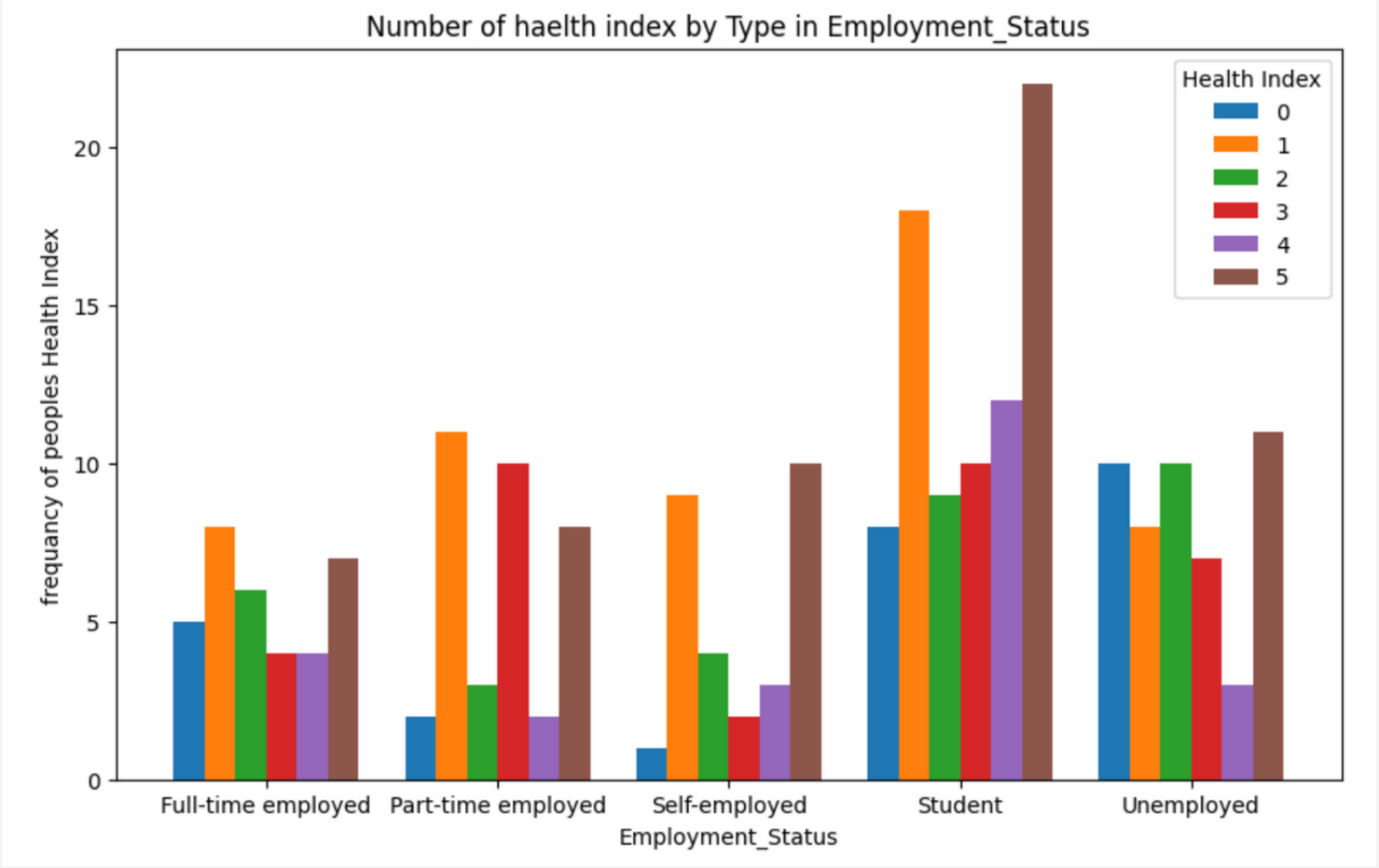
REGENERAL UNDERSTANDINGS





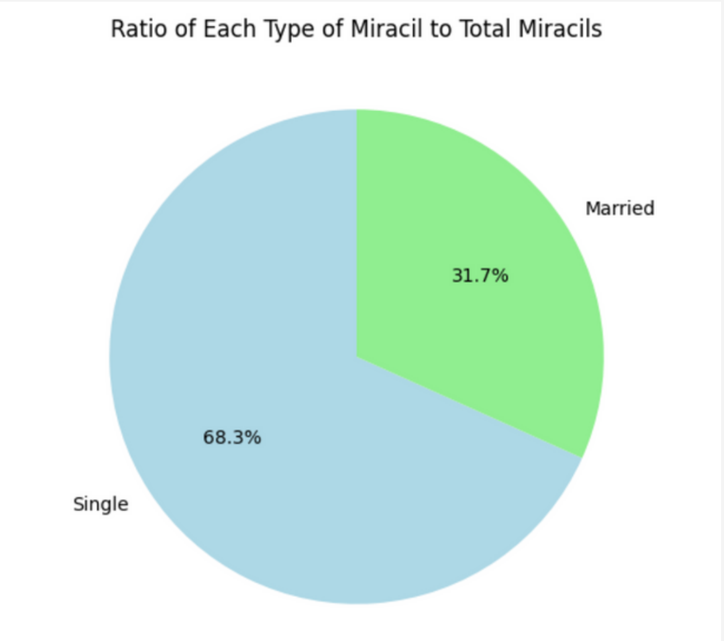
RESULTS AND INSIGHTS

Health Index	0	1	2	3	4	5
Employment_Status						
Full-time employed	5	8	6	4	4	7
Part-time employed	2	11	3	10	2	8
Self-employed	1	9	4	2	3	10
Student	8	18	9	10	12	22
Unemployed	10	8	10	7	3	11



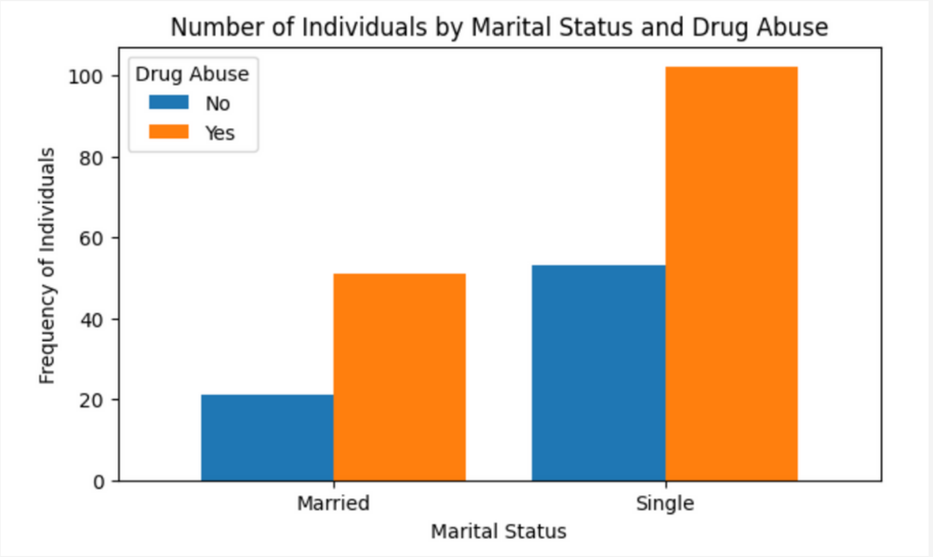
RESULTS AND INSIGHTS

- the data shows that there are more single individuals in the dataset than married ones. This could mean that single people are more common in the dataset, or there might be reasons why more single people are included compared to married ones.
- so the question is: are the single people more open to have drugs?



- full-time and married, are there specific stressors or workplace conditions that could be influencing this trend?
- Are there differences in job stability or financial security between married and single part-time workers that could influence drug abuse behaviors?
- Students: Are there differences in job stability or financial security between married and single part-time workers that could influence drug abuse behaviors?
- Unemployed: How does marital status intersect with unemployment status to influence drug abuse patterns? (marriage point does it effect?)

	Drug_or_not	No	Yes
Employment_Status	**Marital_Status**		
Full-time employed	**Married**	6	19
	Single	4	5
Part-time employed	**Married**	2	9
	Single	8	17
Self-employed	**Married**	12	13
	Single	0	4
Student	**Single**	30	49
Unemployed	**Married**	1	10
	Single	11	27



RESULTS AND INSIGHTS

- full-time and married, are there underlying factors such as job satisfaction or stress levels contributing to these patterns?
- What is the cause of absences?

- Among full-time employed individuals, there are 14 reported cases of drug abuse among those who are married and 4 cases among single individuals.
- Self-employed individuals, married are 9, while singles are 2.
- Students, particularly single ones, exhibited a higher likelihood of drug abuse, with 39 cases reported among single students.
- Unemployed individuals, regardless of marital status, displayed varying degrees of drug abuse, with 7 cases among the married and 18 among the single.

Employment_Status	Full-time employed	Part-time employed	Self-employed	Student	Unemployed	All
Drug_or_not						
No	10	10	12	30	12	74
Yes	24	26	17	49	37	153
All	34	36	29	79	49	227

		Drug_or_not	No	Yes
Employment_Status	Marital_Status	absence_from_work		
Full-time employed	Married	No	6	5
		Yes	0	14
	Single	No	4	1
		Yes	0	4
Part-time employed	Married	No	2	0
		Yes	0	9
	Single	No	7	4
		Yes	1	13
Self-employed	Married	No	12	4
		Yes	0	9
	Single	No	0	2
		Yes	0	2
Student	Single	No	28	10
		Yes	2	39
Unemployed	Married	No	0	3
		Yes	1	7
	Single	No	11	9
		Yes	0	18

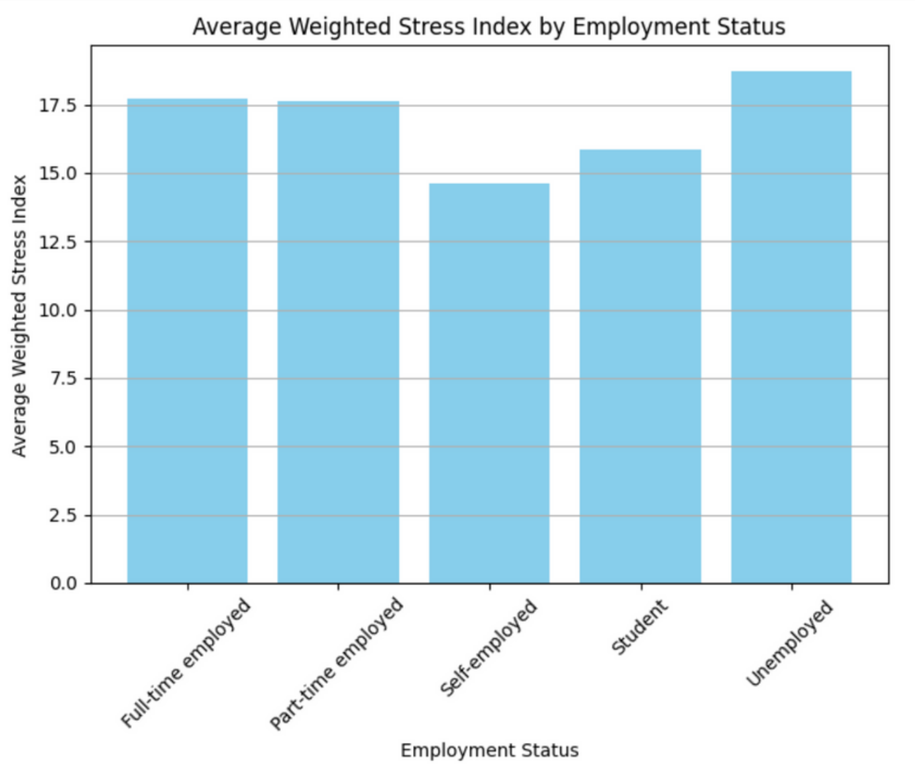
RESULTS AND INSIGHTS

how the average Weighted Stress Index varies across different employment statuses:

- Somehow the stress average of full-time and part-time are near to each other.
- self-employed have the least average

	Employment_Status	Weighted_Stress_Index
0	Full-time employed	17.727034
1	Part-time employed	17.602960
2	Self-employed	14.630818
3	Student	15.837807
4	Unemployed	18.723158

-What factors contribute to the differences in stress levels among various employment statuses?



RESULTS AND INSIGHTS

Heatmap for all the features, open jupyter notebook

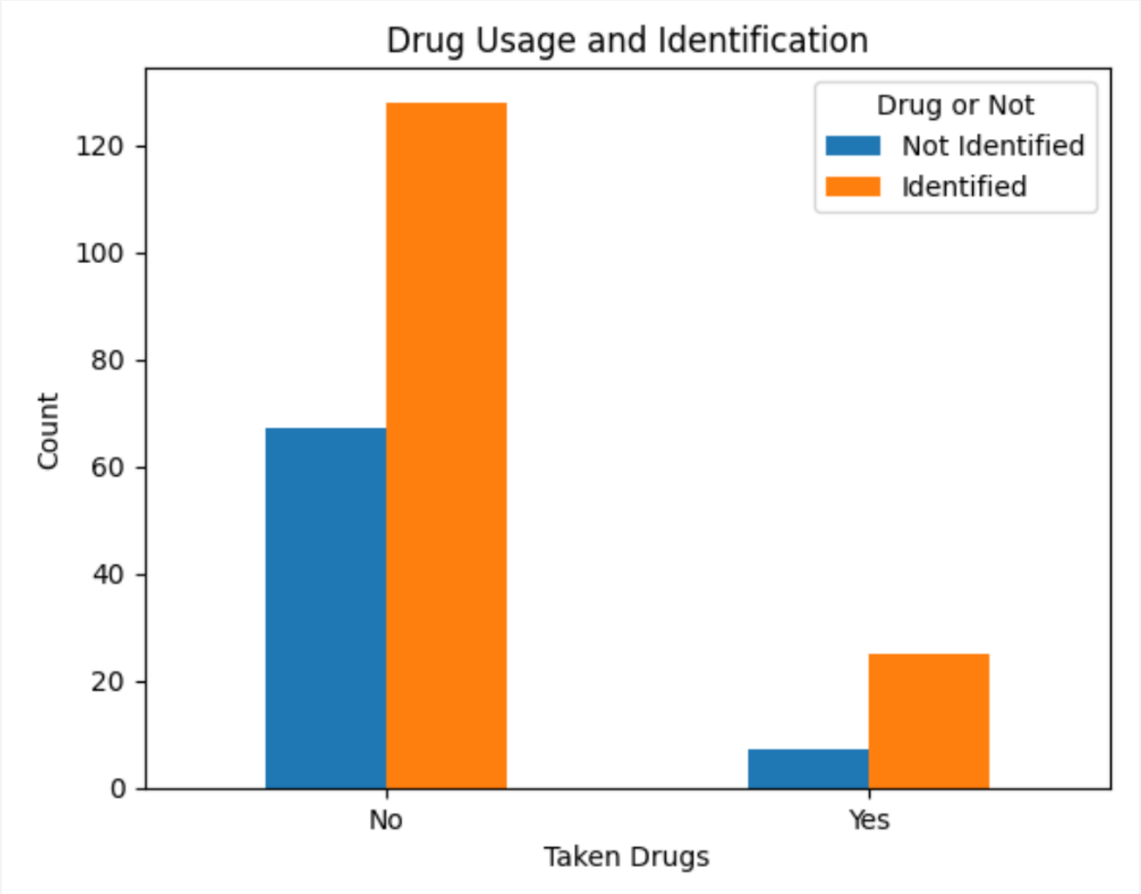
- smoking has a strong positive correlation with taking drugs, absence from work, absent-minded, sleeping habits, and decreased social interaction, lying. This indicates that individuals with no smoking tend to have fewer instances of drug use, fewer absences from work, better sleeping habits, and fewer car accidents.
- taken drugs has a positive correlation with absence from work, indicating that individuals with a history of drug use are more likely to have absences from work, with a bad sleeping habit, and increased personal expenses.
- decreased social interaction has a positive correlation with increased personal expenses, weight changed, lying, sleeping habits, indicating that individuals who decrease their social interactions are more likely to experience increased personal expenses.
- Marital Status has a negative correlation with Income Level, suggesting that individuals with higher income levels are less likely to be single.
- Employment Status has a negative correlation with Income Level, indicating that individuals with higher income levels are less likely to be unemployed or have part-time employment.
- Smoking and Drug or not have a strong positive correlation, indicating that individuals who smoke are more likely to be identified as drug abusers.
- Health Index has strong negative correlations with Smoking, taken drugs, absence from work, sleeping habits, and car crash, suggesting that these behaviors are associated with poorer health.

Drug_or_not Smoking	No	Yes	All
No	64	7	71
Yes	10	146	156
All	74	153	227

sleeping_habits Smoking	No	Yes	All
No	66	5	71
Yes	23	133	156
All	89	138	227

RESULTS AND INSIGHTS

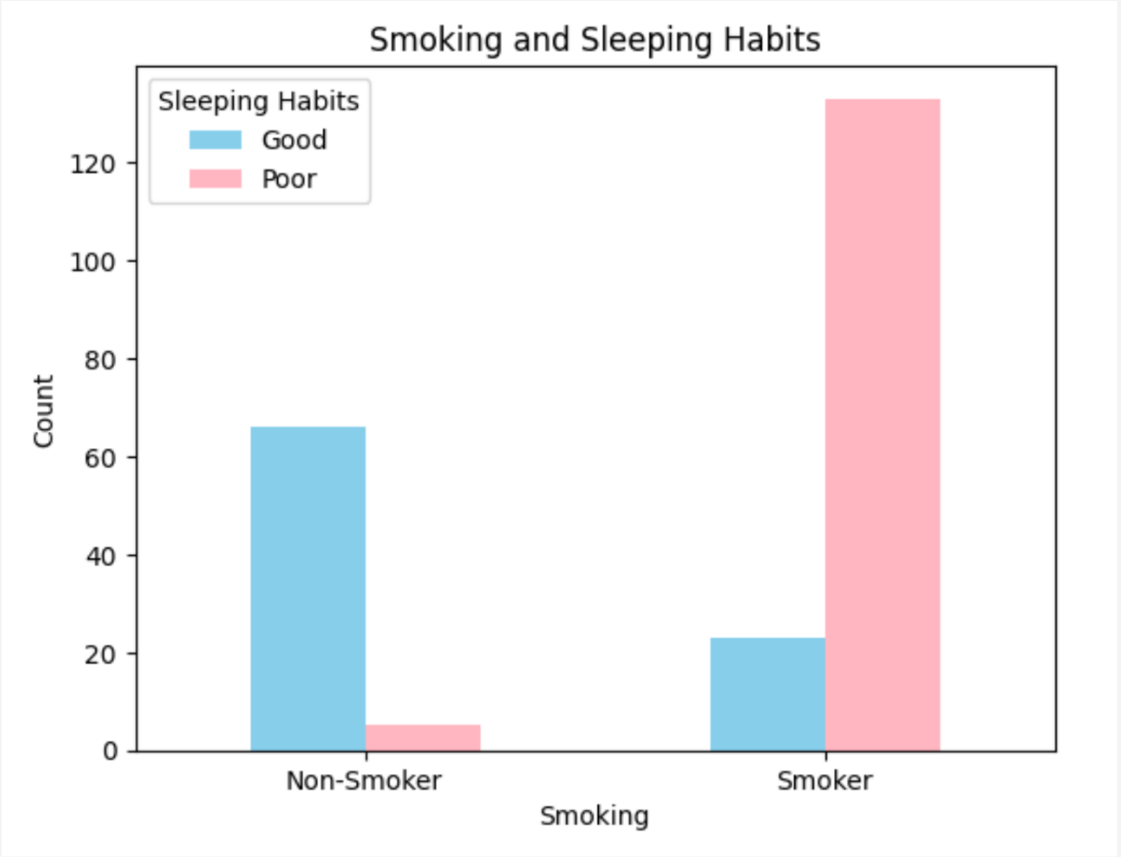
- For those who haven't taken drugs, most haven't been identified as drug abusers.
- Among those who have taken drugs, most have been identified as drug abusers.
- If someone has taken drugs, they're more likely to be identified as a drug abuser.
- Those who haven't been identified as drug abusers might not exhibit typical signs or haven't been caught.



RESULTS AND INSIGHTS

- Among individuals who do not smoke, the majority (93.75%) do not take drugs and have good sleeping habits, while a small proportion (6.25%) do not take drugs and have poor sleeping habits.
- For individuals who smoke, a large proportion (85.71%) who do not take drugs have good sleeping habits, while a smaller proportion (14.29%) have poor sleeping habits.
- Among individuals who smoke and do not take drugs, the frequency of both good and poor sleeping habits is equal (50% each).
- The majority (87.67%) of individuals who smoke and take drugs have poor sleeping habits, while a smaller proportion (12.33%) have good sleeping habits.

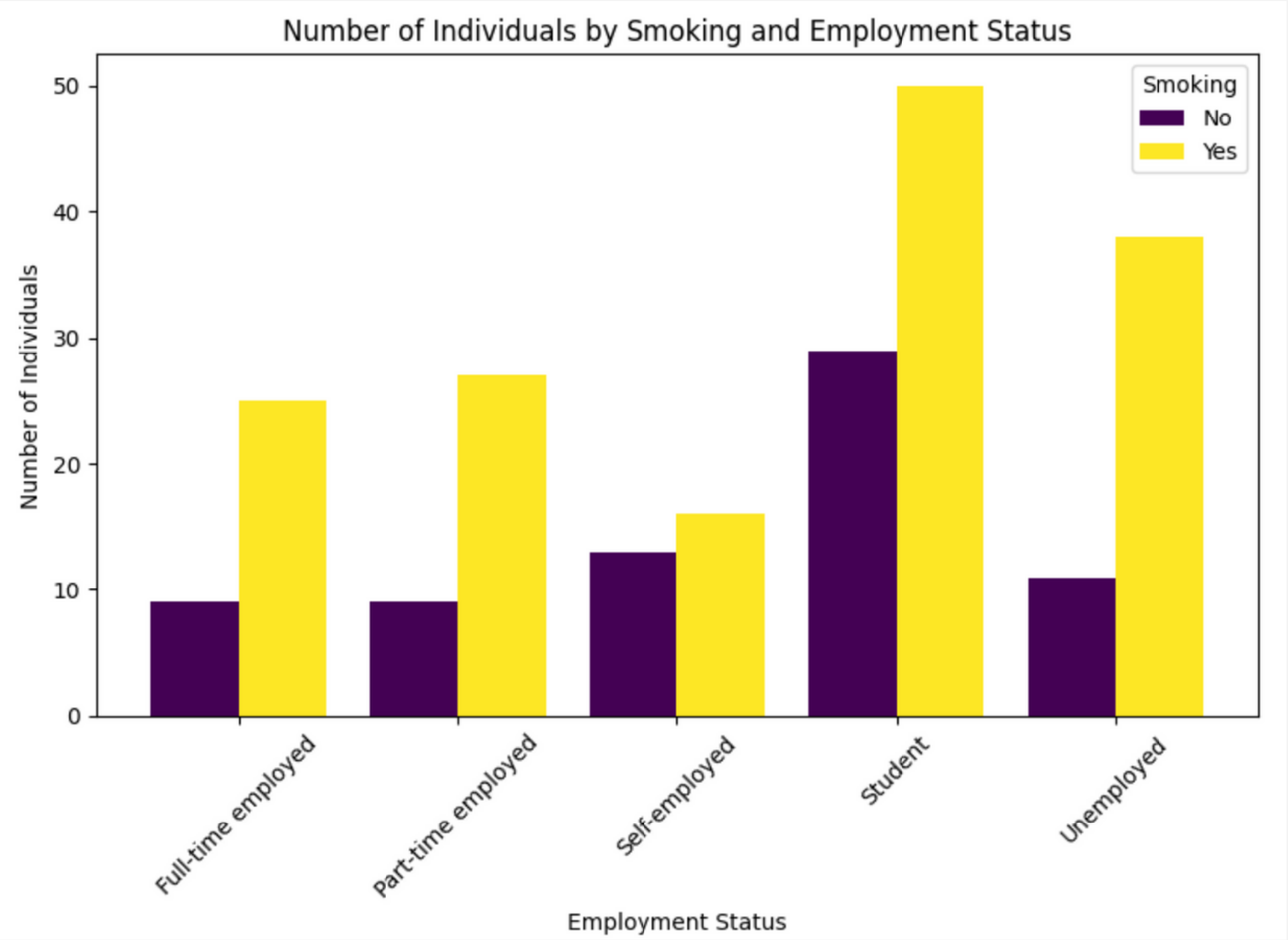
	sleeping_habits	No	Yes
Smoking	**Drug_or_not**		
No	**No**	0.937500	0.062500
	Yes	0.857143	0.142857
Yes	**No**	0.500000	0.500000
	Yes	0.123288	0.876712



RESULTS AND INSIGHTS

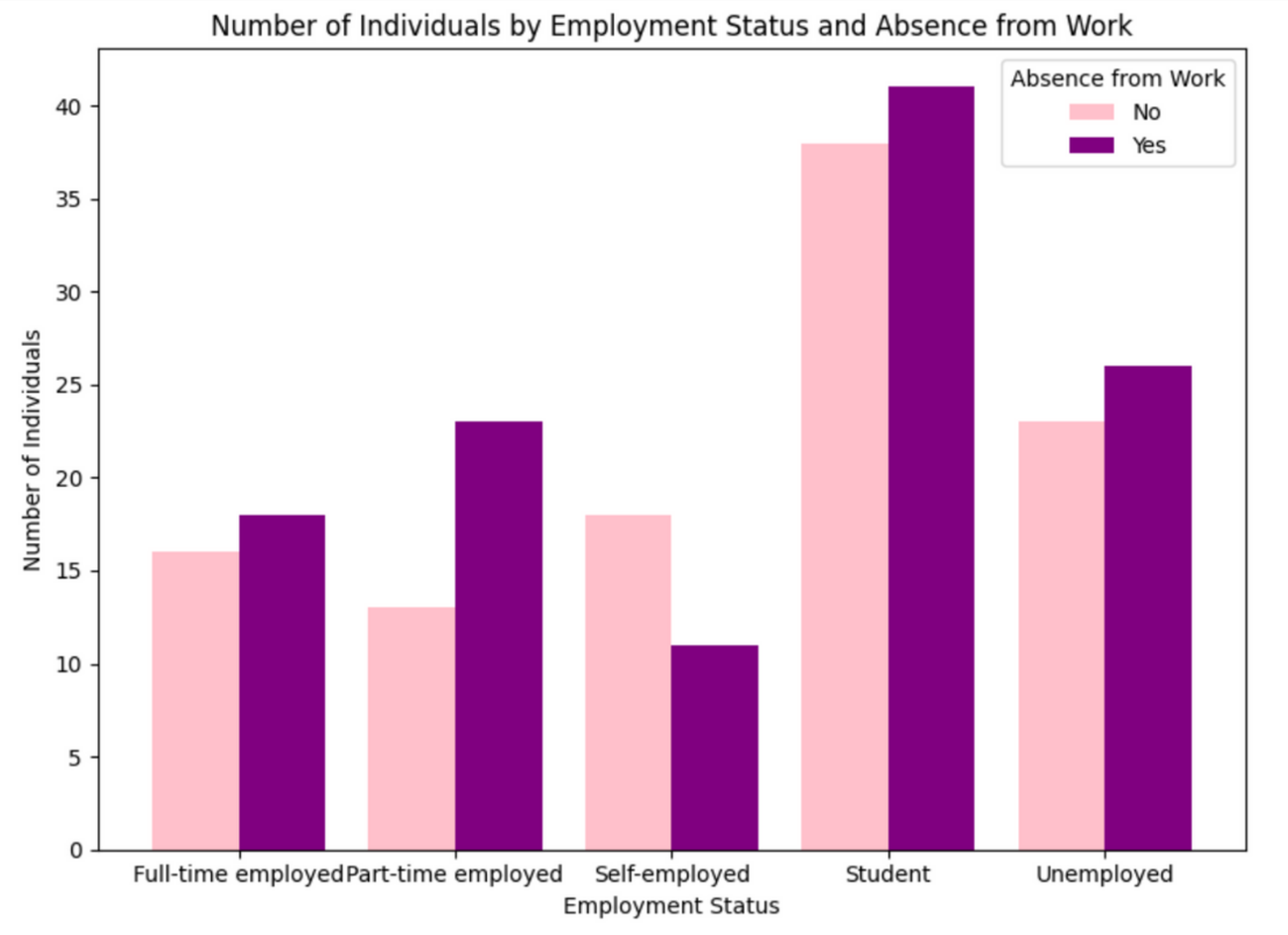
Does employment status influence smoking habits?
there appears to be an association between employment status and smoking habits. For instance, the smoking prevalence is higher among students (50.0%) and unemployed individuals (38.0%) compared to full-time employed (25.0%) and part-time employed individuals (27.0%)

there are notable differences in smoking habits between students and employed individuals. Students exhibit a higher smoking prevalence (50.0%) compared to full-time employed (25.0%) and part-time employed individuals (27.0%).



RESULTS AND INSIGHTS

absence_from_work	No	Yes	All
Employment_Status			
Full-time employed	16	18	34
Part-time employed	13	23	36
Self-employed	18	11	29
Student	38	41	79
Unemployed	23	26	49
All	108	119	227

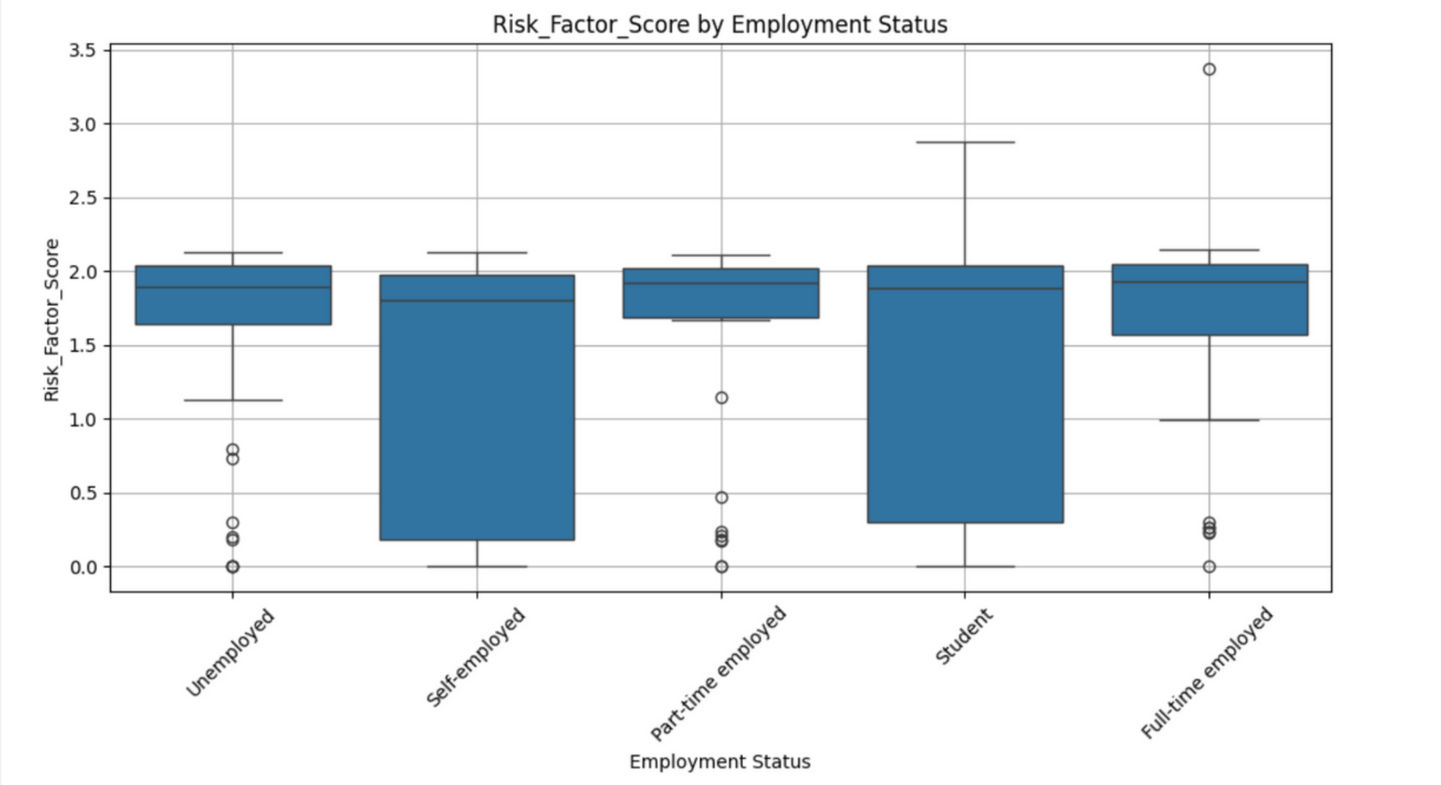


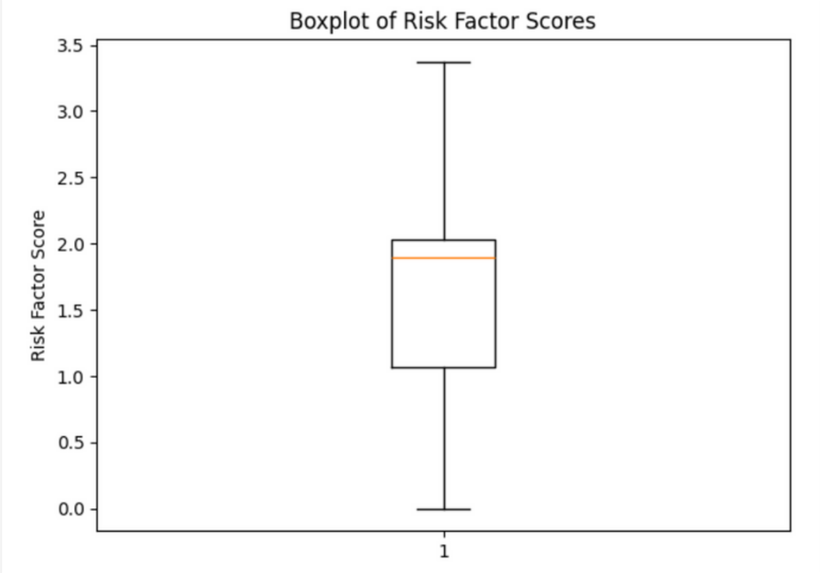
RESULTS AND INSIGHTS

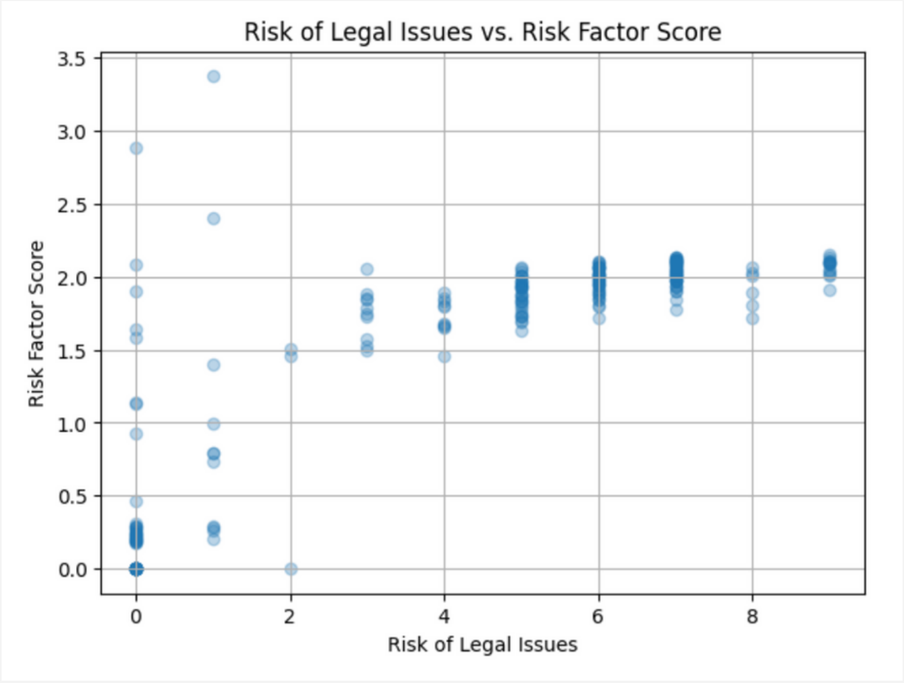
		Employment_Status	Full-time employed	Part-time employed	Self-employed	Student	Unemployed
absence_from_work	Smoking	Drug_or_not					
No	No	No	8	7	11	26	10
		Yes	1	1	2	0	0
	Yes	No	2	2	1	2	1
		Yes	5	3	4	10	12
Yes	No	No	0	0	0	2	0
		Yes	0	1	0	1	1
	Yes	No	0	1	0	0	1
		Yes	18	21	11	38	24

RESULTS AND INSIGHTS

- positive relationship between them, as one increases the other increase too. because the correlation between them is 0.82








```
Correlation coefficient between Risk of Legal Issues and Risk Factor Score: 0.8227700380035544
```

RESULTS AND INSIGHTS

Drug_or_not lying	No	Yes	All
No	73	43	116
Yes	1	110	111
All	74	153	227

Drug_or_not weight_changed	No	Yes	All
No	58	32	90
Yes	16	121	137
All	74	153	227

Drug_or_not selling_house_content	No	Yes	All
No	73	92	165
Yes	1	61	62
All	74	153	227

Drug_or_not commit_stole	No	Yes	All
No	73	102	175
Yes	1	51	52
All	74	153	227

Drug_or_not distracted_and_absent_minded	No	Yes	All
No	62	36	98
Yes	12	117	129
All	74	153	227

Drug_or_not decreased_social_interaction	No	Yes	All
No	68	39	107
Yes	6	114	120
All	74	153	227

D E C I S I O N S

High possibility of having drugs:

- student, smoking, lying, weight changes, distracted and absent-minded, decreased social interaction, increased personal expenses.
- full-time employed, married, absence from work, decreased social interaction, increased personal expenses.
- unemployed, single, smoking, lying, sleeping habits, absence from work, decreased social interaction, increased personal expenses.
- part-time employed, single, smoking, lying, sleeping habits, absence from work, decreased social interaction, increased personal expenses, weight changes.
- part-time employed, married, smoking, lying, sleeping habits, absence from work, decreased social interaction, increased personal expenses, weight changes.
- self-employed, married, smoking, lying, sleeping habits, absence from work, decreased social interaction, increased personal expenses, weight changes.
- unempolyed, single, decreased social interaction, increased personal expenses.

D E C I S I O N S

moderate likelihood of having drugs:

- student, absence from work, increased personal expenses.
- full-time employed, married, absence from work, decreased social interaction.
- self-employed, married.
- unemployed, absence from work, decreased social interaction.

low likelihood of having drugs:

- self-employed, married.
- students.
- unemployed, single.

EVALUATE THE IMPORTANCE OF DATA ANALYTICAL TECHNIQUES

- The tools of the measures of frequency were used for getting an overview of the data (columns), for example, It shows that only 31.5% of people are married and 68.5% of people are single in the entire population. This suggests there are more single people in the dataset than married people (and then when making a visualization for the employment status we will be noticing that the students are the most frequent so it justifies that the single ones are the most frequent). Moreover, of the people who are examined for drug abuse behaviors, the most frequent of the single people are found to be drug abusers (students), while the married people are with a low percentage. This suggests that drug abuse is more common in single people (students) than in married people.

- The central measures of numerical data can be understood through measures of central tendency, such as mean and median. For instance, the average Weighted Stress Index for people who work full-time is 17.72 whereas it is 17.6 for people who work part-time. This suggests that, on average, people who work part-time have slightly higher stress levels than people who work full-time even though it's not that difference. self-employed have the least average.

EVALUATE THE IMPORTANCE OF DATA ANALYTICAL TECHNIQUES

- Data has been arranged in contingency tables to illustrate the relationship between two categorical variables. For example, a contingency table that compares marital status and employment status shows that, among people who work full-time, 73.0% are married and 26.4% are single. Of the unemployed, 77.6% are single and 22.4% are married. This draws attention to the variations in marital status among various job statuses.

- Techniques for detecting outliers have found odd observations that need more research. An outlier analysis, for instance, might show that a particular person has a remarkably high Risk of Legal Issues score in comparison to other people in the dataset, suggesting a possible anomaly that needs more research. for example, unemployed people have many outliers between (0 and 0.8) and this lead us to investigate what aspects and reasons that they are singles, do not smoke etc...

Marital_Status	Married	Single
Employment_Status		
Full-time employed	0.735294	0.264706
Part-time employed	0.305556	0.694444
Self-employed	0.862069	0.137931
Student	0.000000	1.000000
Unemployed	0.224490	0.775510