Project 2

Exploratory Data Analysis

Performing EDA on “drug use by age” data

- By Carol Cheng

GOALS & TASKS - The drug use by age dataset

01

Load the dataset and clean data

- Load the data
- Check data types
- Check if any missing values
- Data cleaning

02

Do a high-level, initial overview of the data

- Use .describe() to check statistics
- Generate some subsets
- Use pd.plot() to visualize data
- Use pd.corr() and sns.heatmap() to explore the correlationships

03

Create a testable hypothesis about the data

- Write a specific question
- Description of the deliverables
- Use EDA techniques to analyse
- Write a report

Prepared the dataset (Data Cleaning)

-Data type

Change the “scope” values in “age” column to single number values (mean), and changed the data type to int

- Null values

Replace ‘-’ with 0 for those missing values. And changed the data type to float

- Set index

Set column ‘age’ as index to facilitate analysis and visualization

- Subset

Create some data subsets for analysis purpose

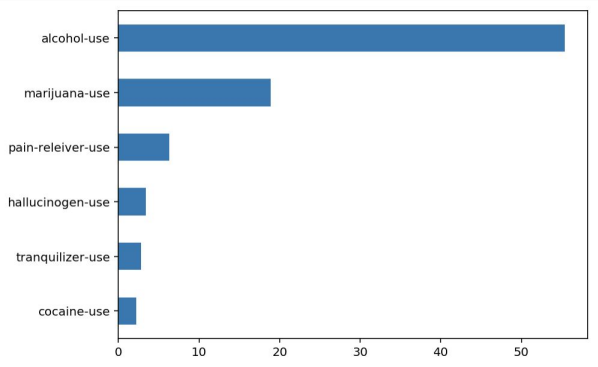
Prepared the dataset (Data Cleaning) - continue

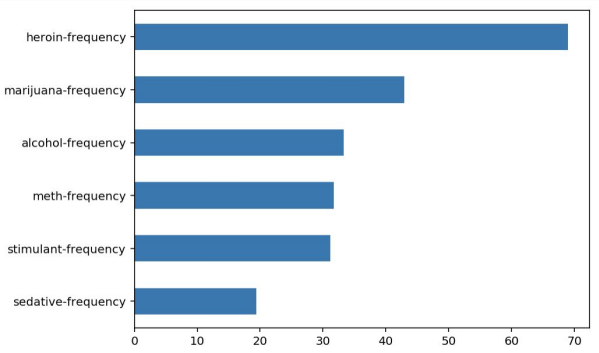
age	n	alcohol-use	alcohol-frequency	marijuana-use	marijuana-frequency	cocaine-use	cocaine-frequency	crack-use	crack-frequency	heroin-use	...
12	2798	3.9	3.0	1.1	4.0	0.1	5.0	0.0	0.0	0.1	...
13	2757	8.5	6.0	3.4	15.0	0.1	1.0	0.0	3.0	0.0	...
14	2792	18.1	5.0	8.7	24.0	0.1	5.5	0.0	0.0	0.1	...
15	2956	29.2	6.0	14.5	25.0	0.5	4.0	0.1	9.5	0.2	...
16	3058	40.1	10.0	22.5	30.0	1.0	7.0	0.0	1.0	0.1	...
17	3038	49.3	13.0	28.0	36.0	2.0	5.0	0.1	21.0	0.1	...
18	2469	58.7	24.0	33.7	52.0	3.2	5.0	0.4	10.0	0.4	...
19	2223	64.6	36.0	33.4	60.0	4.1	5.5	0.5	2.0	0.5	...
20	2271	69.7	48.0	34.0	60.0	4.9	8.0	0.6	5.0	0.9	...
21	2354	83.2	52.0	33.0	52.0	4.8	5.0	0.5	17.0	0.6	...
22	4707	84.2	52.0	28.4	52.0	4.5	5.0	0.5	5.0	1.1	...
24	4591	83.1	52.0	24.9	60.0	4.0	6.0	0.5	6.0	0.7	...
28	2628	80.7	52.0	20.8	52.0	3.2	5.0	0.4	6.0	0.6	...
32	2864	77.5	52.0	16.4	72.0	2.1	8.0	0.5	15.0	0.4	...
42	7391	75.0	52.0	10.4	48.0	1.5	15.0	0.5	48.0	0.1	...
57	3923	67.2	52.0	7.3	52.0	0.9	36.0	0.4	62.0	0.1	...
65	2448	49.3	52.0	1.2	36.0	0.0	0.0	0.0	0.0	0.0	...

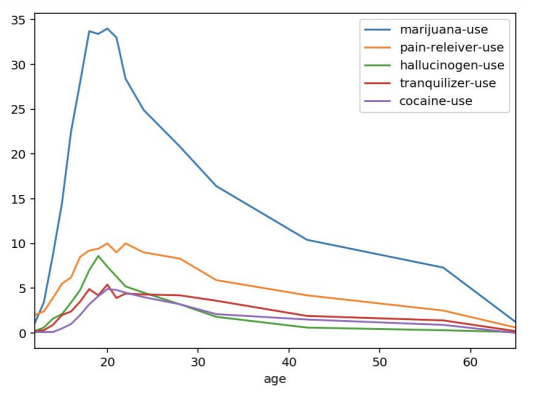
17 rows × 27 columns

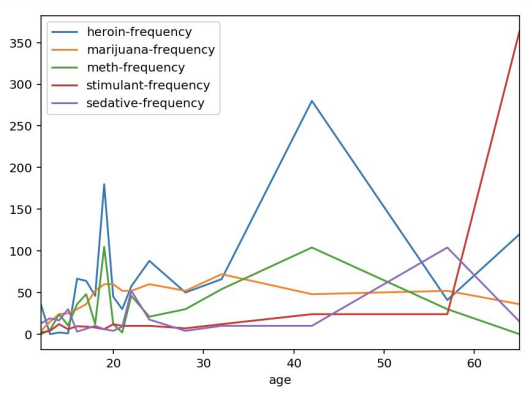
High-level, initial overview of the data

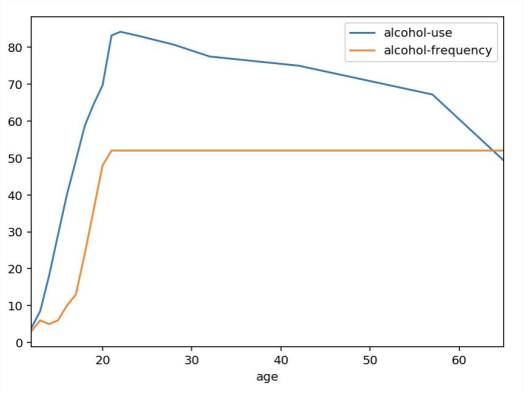
Visualisation





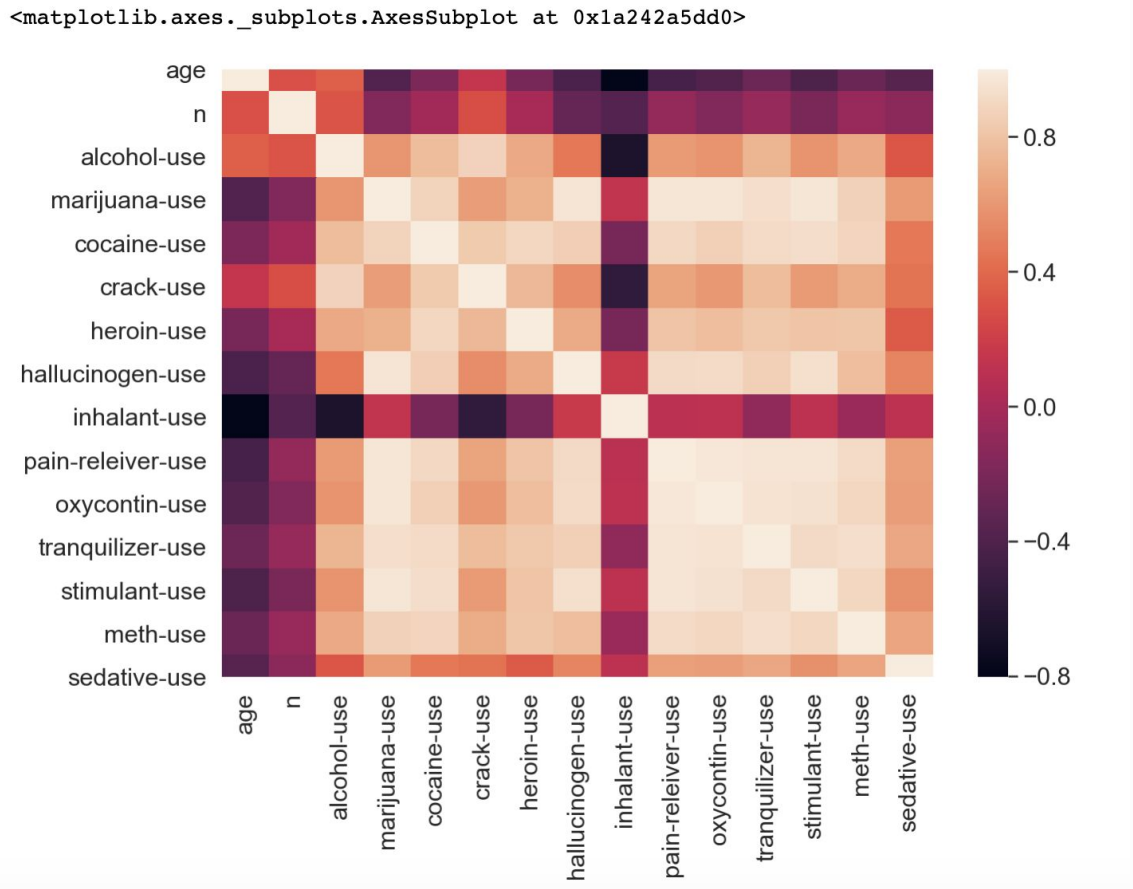






High-level, initial overview of the data - continue

Correlations



Testable Hypothesis

Question

What was the most commonly used prescription drug among all types of prescription drugs?

Background

- Prescription drugs are widely spread all over the world. They can be more easily accessed than hard street drugs.
- Prescription drugs can seem being safer. However, when overdose or misused, they can do great harm to people.
- The previous initial overview of the dataset, it looks there are some strong positive correlations exist between the prescription drugs.
- Aim to find out which drug might have higher risk of causing additive and abuse compare to the other types of prescription drugs.

Hypothesis

Pain-reliever was the most commonly used prescription drug.(There were obvious differences between the use of pain-reliever and the uses of other types of prescription drugs)

Deliverables

- What prescription drug was used by highest percentage of people in the past 12 months?
- How prescription drugs were used among different age groups?
- If an obvious difference exist on drugs use among different age groups?
- Why we should pay attention to prescription drugs abuse? What age groups should we pay more attention on?

Testable Hypothesis

Definetions for this ananlysis purpose:

Age groups:

Teenager: Age under 18
Youth: Age from 18 to 24
Young Adult: Age from 25 to 35
Middle Age: Age from 35 to 64
Senior: equals or older than 65
In this dataset:
An average age '42' represent a sample age group '35-49',
An average age '57' represent a sample age group '50-64',
An average age '65' represent equals or above 65

Pain-reliever

Oxycontin

Tranquilizer

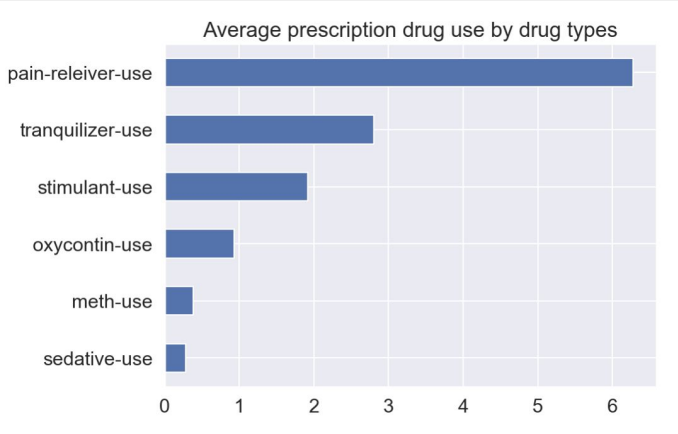
Stimulant

Meth

Sedative

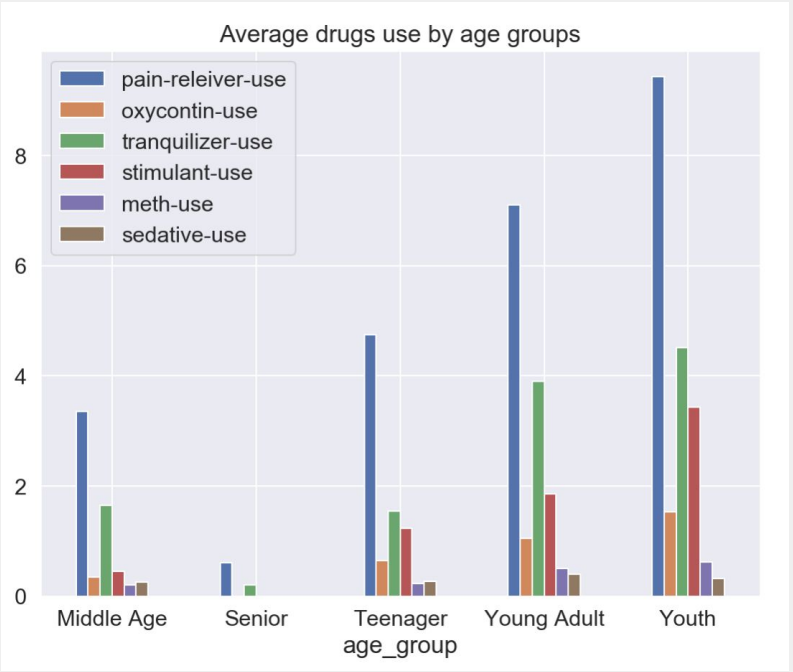
age	pain-releiver-use	oxycontin-use	tranquilizer-use	stimulant-use	meth-use	sedative-use	age_group
12	2.0	0.1	0.2	0.2	0.0	0.2	Teenager
13	2.4	0.1	0.3	0.3	0.1	0.1	Teenager
14	3.9	0.4	0.9	0.8	0.1	0.2	Teenager
15	5.5	0.8	2.0	1.5	0.3	0.4	Teenager
16	6.2	1.1	2.4	1.8	0.3	0.2	Teenager
17	8.5	1.4	3.5	2.8	0.6	0.5	Teenager
18	9.2	1.7	4.9	3.0	0.5	0.4	Youth
19	9.4	1.5	4.2	3.3	0.4	0.3	Youth
20	10.0	1.7	5.4	4.0	0.9	0.5	Youth
21	9.0	1.3	3.9	4.1	0.6	0.3	Youth
22	10.0	1.7	4.4	3.6	0.6	0.2	Youth
24	9.0	1.3	4.3	2.6	0.7	0.2	Youth
28	8.3	1.2	4.2	2.3	0.6	0.4	Young Adult
32	5.9	0.9	3.6	1.4	0.4	0.4	Young Adult
42	4.2	0.3	1.9	0.6	0.2	0.3	Middle Age
57	2.5	0.4	1.4	0.3	0.2	0.2	Middle Age
65	0.6	0.0	0.2	0.0	0.0	0.0	Senior

Testable Hypothesis




```
# Check the average frequency of drug use by different types of drugs
P_drugs_grouped_frequency.mean()
```

```
pain-releiver-frequency     14.705882
oxycontin-frequency         13.941176
tranquilizer-frequency      11.735294
stimulant-frequency         31.147059
meth-frequency              31.735294
sedative-frequency          19.382353
dtype: float64
```



Finding

- Pain-reliever was much more commonly used by people than other types of prescription drugs in this dataset.
- In each age group, the percentage of pain-reliever use was much higher than the others without exception.
- The average frequency of using pain-reliever was less than other drugs like stimulant, meth and sedative.
- The overall prescription drugs use in young people (age under 35) was much higher than older people (age above 35), which indicates young people more relied on pain-reliever than older people

Conclusion

Pain-reliever was the most commonly used prescription drugs for people across all age groups. Although it appears to have lower risk of causing addiction compared to some other types of prescription drugs, pain-reliever drug abuse can still be a severe issue. Further, we should pay attention to pain-reliever use among young people, because base on the dataset, young people were much more inclined to use pain-reliever than older people (age above 35).

THANKS!

Q & A

carolcheng.au@gmail.com