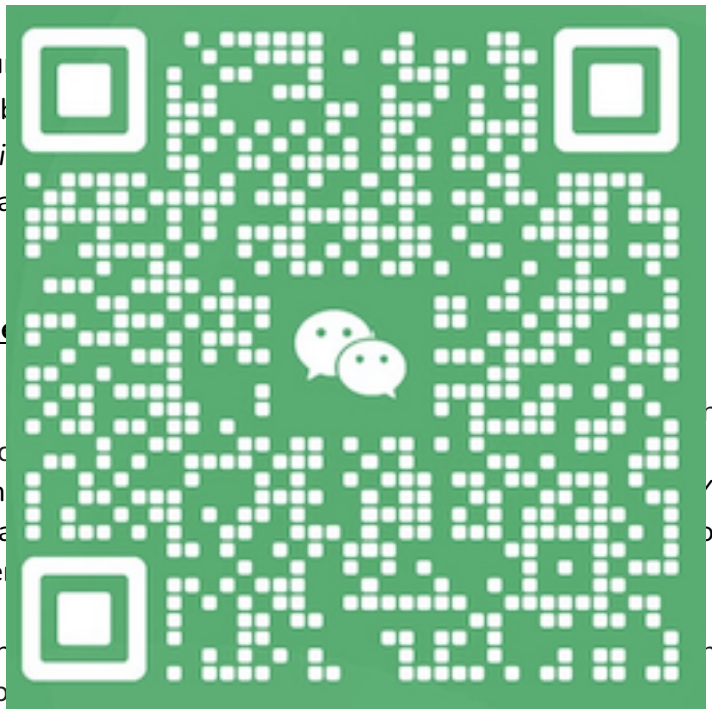
Homework 6 : Analyze Celebrity Deaths

(Deadline as per Canvas)

This homework deals with the following topics:

- The *pandas* module
- Loading data
- Joining data
- Querying data
- Summarizing
- Aggregate fu
- The *numpy* li
- The *matplotli*
- Data visualiza



General Idea of th

In this assignment, ns_2016.xlsx" which contains records of c You'll use functions from the *pandas* m You are expected to summarize data, crea visualize data using histograms and othe

For each question, th nstructions and write the code after each b

```
# your code here
```

Please use the exact variable name if it is specified in the comment.

About the Data

All of the data is contained within the “celebrity_deaths_2016.xlsx” file which contains 2 sheets:

- "celeb_death": contains records of deaths of famous humans and non-humans
    - There are 5 columns: date_of_death, name, age, bio, cause_id

	date of death	name	age	bio	cause_id
2	2016-01-01	Tony Lane	71	American art director (Rolling Stone)	8915
3	2016-01-01	Gilbert Kaplan	74	American conductor and businessman	2860
4	2016-01-01	Brian Johns	79	Australian company director, managing director of the Australian Broadcasting C	2860
5	2016-01-01	Natasha Aguilar	45	Costa Rican swimmer, silver and bronze medalist at the 1987 Pan American Gan	33897
6	2016-01-01	Fazu Aliyeva	83	Russian Avar poet and journalist	10648
7	2016-01-01	Mike Oxle		Representatives fror	7674

- "cause_of_de
    - There

	cause_id	cause_
2	753	ALS
3	1039	bomb
4	1120	shot
5	1499	fall
6	1629	shot
7	2132	gored
8	2151	tased

During this exercise, “cause_of_death” data using the “cause ath.

Other information ab

- The cause of death was not reported for all individuals
- The dataset might include deaths that took place in other years (you'll need to ignore these records)
- The dataset might contain duplicate records (you'll need to remove them)

Submission

To complete the assignment, download *celebrity_deaths_2016.ipynb* and *celebrity_deaths_2016.xlsx*.

Evaluation

Two points for each question.