DS Exam 4 – Pandas

- Answer all the questions below.
- No material / computers are allowed.
- Exam time: 1.5 hours.
- Good luck!

Question 1

What will be the result of the following expressions?

```
# 1
oil_prices = pd.DataFrame({
    'U.K. Brent': {'2013-Q1': 112.9, '2013-Q2': 103.0,
                   '2013-Q3': 110.1, '2013-Q4': 109.4},
    'Dubai':      {'2013-Q1': 108.1, '2013-Q2': 100.8,
                   '2013-Q3': 106.1, '2013-Q4': 106.7},
    'West Texas': {'2013-Q1': 94.4,  '2013-Q2': 94.2,
                   '2013-Q3': 105.8, '2013-Q4': 97.4}})
oil_prices
```

```
# 2
oil_prices.loc['2013-Q2':'2013-Q3', ['West Texas', 'U.K. Brent']]
```

```
# 3
oil_prices['Dubai'][0]
```

```
# 4
oil_prices.loc[:, oil_prices.loc['2013-Q1'] > 110]
```

```
# 5
oil_prices.iloc[0, 1:3]
```

```
# 6
oil_prices.idxmax()
```

```
# 7
oil_prices.unstack()
```

```
# 8
oil_prices.stack()
```

```
# 9
oil_prices.reset_index().iloc[:, 0].str.split('-', expand=True)
```

```
# 10
pd.concat([oil_prices, pd.DataFrame({
                        'Dubai':      {'2014-Q1': 107.3},
                        'U.K. Brent': {'2014-Q1': 102.5},
                        'Urals':      {'2014-Q1': 99.8}})])
```

Question 2

Answer the following questions regarding the flights dataset:

```
flights = pd.read_csv('flights.csv')
flights.head()
```

	Month	Day	Weekday	Airline	Origin	Dest	AirTime	Distance	ArrivalDelay	Diverted	Cancelled
0	1	1	4	WN	LAX	SLC	94.0	590	65.0	0	0
1	1	1	4	UA	DEN	IAD	154.0	1452	-13.0	0	0
2	1	1	4	MQ	DFW	VPS	85.0	641	35.0	0	0
3	1	1	4	AA	DFW	DCA	126.0	1192	-7.0	0	0
4	1	1	4	WN	LAX	MCI	166.0	1363	39.0	0	0

1. Find the month and day in which the longest arrival delay occurred.

2. Find the three airlines with the longest arrival delays.

3. Find the three airlines with the highest number of cancelled or diverted flights.

4. Find the airlines with more than 2% of cancelled flights. For each such airline, print its name and the precentage of cancelled flights.

5. For each origin and destination, find the average and variance of the airtime.

6. Find the airline that makes the most flights between 500 and 1,000 miles.

7. Create a bar plot showing the number of flights originating from each city.

8. Change the Weekday column to have the day name instead of day number, e.g. 1 - Sun, 2 - Mon, etc.

9. Join the Month and Day columns to a single column named Date. For example, if month=3, day=15, the date column should have 3/15.

10. Find the longest sequence of on-time flights per airline (an on-time flight is a flight with less than 15 minutes arrival delay).