Practice Interview Questions

# Data Wrangling

## Pandas

* What data structures does pandas introduce which aren’t native to Python?

DataFrame

* Video: How can you deal with missing values?

1.leave as-is

2.Drop them using .dropna( )

3.Fill them using .fillna( ) with provided value such as ‘missing’

or with summary statistic such as mean or medium,

* What is the difference between the .loc and the .iloc indexers?

loc gets rows (or columns) with particular labels from the index.

iloc gets rows (or columns) at particular positions in the index (so it only takes integers).

* What file formats for storing data do you know?

Txt, excel, csv, json, xml

* What is the standard way of marking missing values in pandas?

use two already-existing Python null values: the special floating-point NaN value, and the Python None object

* Video: What features of pandas do you like particularly? Any that you dislike?

 Apply is a pretty flexible function, it applies a function along any axis of the DataFrame. I will be using [apply](http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.apply.html?highlight=apply#pandas.DataFrame.apply)function to divide each value of the acid by 100.

df = olive\_oil[list\_of\_acids].apply (lambda x: x/100.00)

Dislike

No true query processing layer

Such as filter, group, join, aggregate…

* What kind of indexes exist in pandas DataFrames?

.loc[ ] is primarily label based

.iloc[ ] is primarily integer position based

## SQL

* What are aggregations in SQL?

An aggregate function allows you to perform a calculation on a set of values to return a single scalar value.

* Can you explain the different types of SQL JOINs?

(INNER) JOIN: Returns records that have matching values in both tables

LEFT (OUTER) JOIN: Return all records from the left table, and the matched records from the right table

RIGHT (OUTER) JOIN: Return all records from the right table, and the matched records from the left table

FULL (OUTER) JOIN: Return all records when there is a match in either left or right table

* Give an example of some aggregation functions in SQL.

AVG – calculates the average of a set of values.

COUNT – counts rows in a specified table or view.

* Can you explain the difference between the WHERE and HAVING filters?

WHERE Clause helps to filter out records based on the data available in a Database table.

E.G. Select Population from state where city = ‘NEW YORK’.

HAVING Clause on the other hand helps you to filter records on the basis of results of Aggregation Functions (i.e. Group By clause)

E.G. Select State and Total Population from state where sum(population) is greater than 10000.