Practice Interview Questions

Data Wrangling

|  |
| --- |
| Pandas |
| * What data structures does pandas introduce which aren’t native to Python?   pandas introduces two new data structures to Python - [**Series**](http://pandas.pydata.org/pandas-docs/dev/dsintro.html#series) and [**DataFrame**](http://pandas.pydata.org/pandas-docs/dev/dsintro.html#dataframe), both of which are built on top of [**NumPy**](http://www.numpy.org/) (this means it's fast). Extra: What is it Pandas? [**pandas**](http://pandas.pydata.org/) is an open source [**Python**](http://www.python.org/) library for data analysis. Python has always been great for prepping and munging data, but it's never been great for analysis - you'd usually end up using [**R**](http://www.r-project.org/) or loading it into a database and using SQL (or worse, Excel). pandas makes Python great for analysis. |
| * **Video:** How can you deal with missing values?   several options we have for handling missing values.   1. DROP MISSING VALUES 2. FILL MISSING VALUES WITH TEST STATISTIC 3. PREDICT MISSING VALUE WITH A MACHINE LEARNING ALGORITHM   https://cdn-images-1.medium.com/max/800/1*_RA3mCS30Pr0vUxbp25Yxw.png |
| * What is the difference between the .loc and the .iloc indexers? * .loc is primarily label based, but may also be used with a boolean array. .loc will raise KeyError when the items are not found. * .iloc is primarily integer position based (from 0 to length-1 of the axis), but may also be used with a boolean array. .iloc will raise IndexError if a requested in |
| * What file formats for storing data do you know?   JSON  CSV  TSV  Spreadsheets  Images  XML  HTML |
| * What is the standard way of marking missing values in pandas?   NA – Not Available  NaN – Not a Number  The choice of using NaN internally to denote missing data was largely for simplicity and performance reasons |
| * **Video:** What features of pandas do you like particularly? Any that you dislike?   The ten functions that do the nearly the same thing. Merge and Join are probably the most irritating examples I run into. |
| * What kind of indexes exist in pandas DataFrames? |
| SQL |
| * What are aggregations in SQL?   Sum / mean / count / min / max |
| * Can you explain the different types of SQL JOINs?   Join = inner join  Full (outer) join  Left (outer) join  Right (outer) join |
| * Give an example of some aggregation functions in SQL.   Count(coulumn\_name) AS …. |
| * Can you explain the difference between the WHERE and HAVING filters?   Since “WHERE” can not be used over aggregated queries, we use “HAVING” |

Büyük [sinema](http://www.cumhuriyet.com.tr/etiket/sinema" \t "_blank) festivalleri kaçınılmaz olarak yeniden yapılanma süreci içindeler. Eski hedeflerin, ideallerin, alışkanlıkların sürdürülemeyeceği ortada. Hızlı ve köktenci bir değişim yaşanıyor. Teknolojik gelişmelerin sinemayı bilgisayarımıza ve cebimize soktuğu, televizyon formatının giderek egemenlik kurduğu bir sunum patlamasına nasıl uyum sağlamalı? Her konu masaya yatırılmış durumda. Kimin için festival? Hangi sanat sineması? Hangi ana akım? Hangi bağımsız sanat çizgisi? Hangi amaçla, hangi jüri? Kimi kime rağmen, hangi nedenlerle, nasıl desteklemeli?...