**Unit 5.1**

Rubric for JSON exercises

<https://docs.google.com/document/d/17CODPE-R47snMo22iPHMh9CCmU3y77zn67SkNOpJMI0/edit>

**Unit 5.2**

Brandon Rhodes – Foundations of Pandas Notes

**Unit 5.3**

**SQL Training:**

Single quotes reference values within a column

< or > operators reference alphabetical order

“” for column name (when necessary i.e. has a space or refers to a function)

Aggregators only aggregate vertically (column), rows can be aggregated by operators

**SQL Case Study:**

A Drop in User Engagement

https://mode.com/sql-tutorial/sql-business-analytics-training

*What are some reasons for the user engagement drop?*

* Did a major client drop stop using the service?
  + No way to tell
* Did a major country or client go on holiday?
  + US was responsible for almost half of the drop
* Was there a patch released that had a bug in the code?
  + No way to tell
* Is the tracking code bugged?
  + Probably not
* Did a marketing event trigger the spike and now some users are no longer utilizing the service?
  + Probably not

*Making a recommendation:*

* All countries are tracked throughout the time period in question
* The US accounts for close to 50% of the drop

*Article on NOSQL*

<https://www.thoughtworks.com/insights/blog/nosql-databases-overview>

The CAP theorem states that if you get a network partition, you have to trade off availability of data versus consistency of data. Durability can also be traded off against latency, particularly if you want to survive failures with replicated data.

Key-Value Databases; Document Databases; Column-family (many columns associated with a row key), Graph databases