Reference:

<http://spark.apache.org/docs/2.1.0/api/python/pyspark.sql.html#pyspark.sql.functions.rand>

<https://towardsdatascience.com/filter-aggregate-and-join-in-pandas-tidyverse-pyspark-and-sql-71d60cbd5330>

<http://www.learnbymarketing.com/1100/pyspark-joins-by-example/>

<http://www.alpha-epsilon.de/cca175/2017/09/13/filter-aggregate-join-rank-and-sort-datasets-spark-python/>

<https://www.edureka.co/blog/spark-sql-tutorial/>

<https://hendra-herviawan.github.io/pyspark-dataframe-row-columns.html>

<https://stackoverflow.com/>

Youtube links:

<https://www.youtube.com/watch?v=WQErwxRTiW0>

<https://www.youtube.com/watch?v=EwAqEZRBQqw>