# HW 1 – Apache Cassandra Executive Summary

1. **Dataset selection:**
2. **DB design:**
3. **Data Analysis:**
4. **Insights & Visualization:**
5. **Challenges:**

* For easier readability and interfaces we decided to integrate all code to one Jupyter notebook. CQL commands were given to the cluster using Cassandra Driver, a python-Cassandra API.
* During data ingestion, our original script has loaded one row at the time in a for loop. As we have a very large number of rows in our dataset (about 270K in the largest table) that took a long time. The solution was to use