Big data platform final project executive summary

Project overview:
We have developed a kubernetes based application which is designed to store tweets in a database and query this database in order to find meaningful insight into the data.

Data description
We have a CSV file with the tweets of the 20 most popular twitter users. The dataset includes fields such as the tweets author, tweets content, and the data when the tweet was published.

K8 cluster overview and rationale
Our project uses a microservices architecture where each component is encapsulated in its own pod. We have the following 3 separate pods:
- A front end which allows the user to enter the tweet id they wish to perform sentiment analysis on.
- A back end which interacts with our database and performs the sentiment analysis.
- A postgres database which stores our data persistently.

The design rationale behind this configuration is to allow for easy scaling of resources based on demand to a specific service. Deploying each component separately allowed us to work better as a team as each team member could focus on their own portion of work without heavily relying on other parts of code. This is because each microservice can be developed independently of the other.

Key insights from the data
After running sentiment analysis on the tweets we found that many of them had negative sentiment.