



Web Scrapping Top Pharma Company Revenue from Wikipedia

Leela Krishna Kosaraju

Northwest Missouri State University

Question

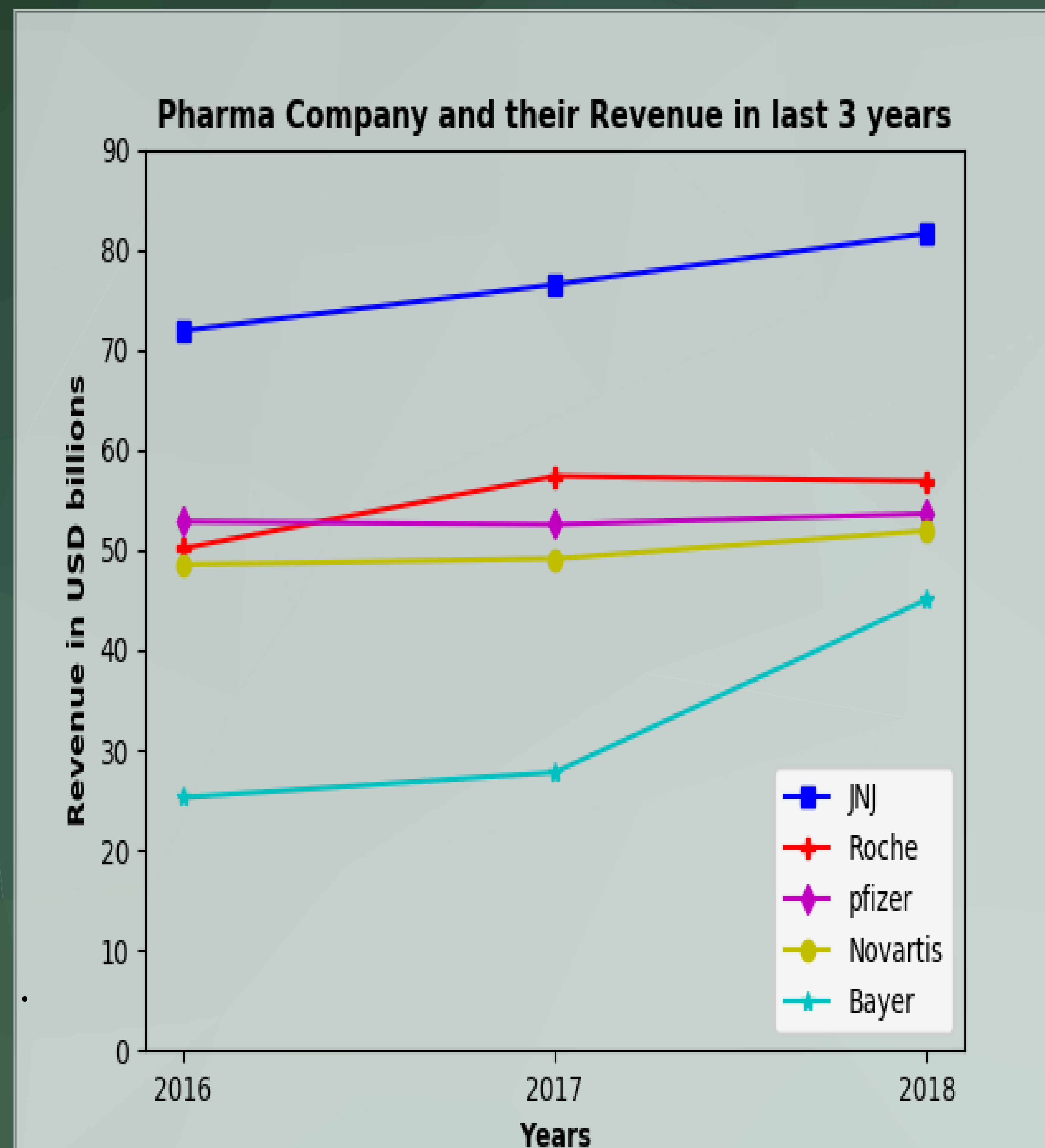
The data of the Top Pharmaceutical companies was taken from Wikipedia. Find out which Pharmaceutical company has the highest Revenue in the last three years.

Methods

This Project used Beautiful Soup, Requests, Pandas, NumPy and Matplotlib libraries for data collection and processing.

First, Wikipedia data is taken and stored in a csv file using Beautiful Soup, Requests and Pandas. Then we use Matplotlib to plot a line graph between pharmaceutical companies and their revenues over the past three years.

Results



The above graph represents the amount of revenue that the pharmaceutical companies earned in the last three years.

Conclusion

From the obtained Line Graph we can say that:

1. Johnson & Johnson(JNJ) Company has the highest revenue in all the three years and steadily increasing
2. Roche has 50 billion dollars revenue in 2016, 14% rise in the following year and maintained the same in 2018
3. Pfizer has almost same revenue in past three years which was around 52billion
4. Novartis has slight increase in their revenue in 2018
5. Even though the Bayer having least revenue among the all in 2016 and 2017. In 2018 this company earned drastic rise in their revenue which was 45 billion dollars

In last three years, Johnson & Johnson has always the highest revenue, and Bayer has always the lowest revenue compared to other pharmaceutical companies.