Q1(a) Uber transforms the traditional business into a digitalized business in several ways. Their approach will be discussed in terms of the 5Vs of big data. First, Uber handles an enormous amount of data every day. There are over a hundred petabytes of data such as Presto queries, Spark jobs and Hive queries in the database: Hadoop Distributed File System, Uber’s compute cluster. [1]

Next, Uber utilises various data in their operation. Its dynamic pricing model uses diverse data such as the location of users, street traffic and demand data. Thus, it adjusts fares in real-time based on the demand and the areas of users. [2] Moreover, Uber Eats and Connect services contain data such as restaurants’ menus and delivery details.

Besides, Uber’s Expected Time of Arrival (ETA), requires high veracity. It uses a Hybrid approach to calculate the ETA: a physical model that uses map data and real-time traffic data between two locations and machine learning to predict the residual between the ETA and the observed outcome. Different spatial and temporal features such as the traffic conditions and the nature of the request are used to train the model. Thus, it increases the performance of the system and benefits all Uber service users. [3]

Also, a rating system to maintain the standard of riders and drivers. If drivers score below a certain threshold, they will not be offered any more work. Users with higher ratings will be prioritised. Therefore, it is valuable to keep riders' and drivers’ experience safe, comfortable and enjoyable, thus, to retain users and maintain profit. [4] [5]

Uber derives value from various data by adjusting fares dynamically to match the supply and demand. It analyses the data during peak hours and historical data such as weather conditions and local events. It maximises Uber’s revenue and ensures that customers can get rides. Also, by utilising real-time traffic data, route optimization reduces travel time and thus improves efficiency.

In the pricing model, the traditional taxi business only relies on a fixed pricing model where Uber utilises the data on riders’ demand and traffic patterns to implement dynamic pricing. For example, finding a taxi is difficult during peak hours and in remote areas. The model can adapt to market fluctuations and customer demand which traditional pricing models cannot.

Moreover, traditional taxi drivers usually rely on their knowledge to drive. Sometimes, taxi drivers intentionally drive longer distances to get more fares, since the routes are not restricted. Nevertheless, Uber’s route optimization can recommend the most efficient routes for drivers to reach their destinations. It reduces traveling costs and the risk of customers paying more due to drivers’ route decisions. This data-driven function maximizes efficiency and minimises transportation costs compared to traditional business.

Q1(b) I will discuss the Netflix industry in the following. First, Netflix is handling a significant amount of data every day. There are around 250 million Netflix users [6] and Netflix processes and analyses vast volumes of data users’ data in real time using Apache Hadoop and Apache Spark. [7]

Also, Netflix utilises various data in its operation. It uses data to enhance personalization, global expansion and original content production. It helps Netflix to retain users and maintain profit.

Moreover, Netflix needs to use data at a high accuracy rate. For example, Netflix should not recommend movies with adult content. Therefore, Netflix designs, implements, and evaluates machine learning models and algorithms through both offline experiments and online A/B testing. It also applies diverse algorithmic approaches including causal modeling, bandits, reinforcement learning, ensembles, neural networks, probabilistic graphical models, and matrix factorization. [8]

Besides, Netflix extracts value from the data especially in its recommendation system. It suggests personalized content to users based on their viewing history, preferences, and social trends. Netflix measures how users respond to the changes in the system by A/B tests and analyses the long-term satisfaction metrics. [9]

Next, Netflix analyses data in real-time and provides immediately updated applications. It uses a real-time analytics database, Apache Druid, which has a high performance at data visibility and operational analytics, and handling, to handle fast queries and high concurrency. [10]

Netflix's recommendation system derives value from the data by providing personalised content, maintaining predictive analytics, curating content, and continuously improving recommendation accuracy. It improves the user experience and thus enhances user satisfaction and retention.

Traditional businesses, such as movie rental stores, typically lack the usage of data to offer insights such as personalized recommendations. They usually only rely on general marketing strategies or fixed sales schedules. However, Netflix applies advanced algorithms and data analytics to provide subscribers with personalised recommendations, improving users’ engagement and satisfaction.

**References**

[1] Reza Reza. (2018, Oct. 17). *Uber’s Big Data Platform: 100+ Petabytes with Minute Latency* [Online]. Available: <https://www.uber.com/en-HK/blog/uber-big-data-platform/>

[2] Yannik Dobler. (2022, Oct. 4). *Uber knows you: how data optimizes our rides* [Online]. Available: <https://d3.harvard.edu/platform-digit/submission/uber-knows-you-how-data-optimizes-our-rides/>

[3] Shloka Daga. (2023, Aug. 9) *How uber uses data? Part 01* [Online]. Available: <https://www.linkedin.com/pulse/how-uber-uses-data-part-01-shloka-daga/>

[4] Bernard Marr. (2015, May. 24) *The Amazing Ways Uber Is Using Big Data Analytics* [Online]. Available:<https://www.linkedin.com/pulse/amazing-ways-uber-using-big-data-analytics-bernard-marr/>

[5] Uber. (2018, Nov. 22) *How the Uber rating system works* [Online]. Available:

<https://www.uber.com/en-EG/blog/how-the-uber-rating-system-works/>

[6] Julia Stoll. (2024, Jan) *Number of Netflix paid subscribers worldwide from 1st quarter 2013 to 3rd quarter 2023* [Online]. Available: <https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/>

[7] VivekR. (2023, Jun. 18) *How Did Netflix Use Big Data to Transform Their Company and Dominate the Streaming Industry?* [Online]. Available:

<https://vivekjadhavr.medium.com/how-did-netflix-use-big-data-to-transform-their-company-and-dominate-the-streaming-industry-a93f90ae8dad>

[8] Netflix Research. *Machine Learning* [Online]. Available:

<https://research.netflix.com/research-area/machine-learning>

[9] Netflix Research. *Recommendations* [Online]. Available:

<https://research.netflix.com/research-area/recommendations>

[10] Ben Sykes. (2020, Mar. 4) *How Netflix uses Druid for Real-time Insights to Ensure a High-Quality Experience* [Online]. Available: <https://netflixtechblog.com/how-netflix-uses-druid-for-real-time-insights-to-ensure-a-high-quality-experience-19e1e8568d06>

一張含有 文字, 螢幕擷取畫面, 字型, 圖表 的圖片

自動產生的描述Q2



From the bar charts, the number of real trees fluctuates around 27 million. There is no significant uprising or downward trend for the number of real trees, possibly because the demand for real trees has reached a saturation point. However, the number of fake trees is rising significantly, from 8.2 million in 2010, an increase of around 127%, to 18.6 million in 2016.

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述(2)

The pie charts show the proportion of fake trees of the entire Christmas trees market is increasing, from around 20% in 2010 to more than 40% in 2016.

Both Figures (1) and (2) can be explained by the increasing popularity of fake trees. It suggests that consumers are more willing to find artificial alternatives since fake Christmas trees are more convenient, cost-effective, and environmentally friendly.

一張含有 行, 文字, 圖表, 繪圖 的圖片

自動產生的描述(3)

First, from the line graphs, the average price of fake trees is lower than that of real trees. The most probable rationale is that the production cost of fake trees is lower. Real trees require more human resources to take care of and more caution while transporting. Moreover, fake tree prices grow quickly in the later years, and its cause may be that society is pursuing a greener environment. Thus, a higher demand will lead to high average fake tree prices.

一張含有 文字, 行, 圖表, 繪圖 的圖片

自動產生的描述(4)

From another line graph plotted by year against sales, the sales of fake and real trees are rising. Figures (3) and (4) show increasing demands and sales of fake and real trees. It may thus indicate that the society is experiencing general economic growth, in which people have more wealth than before, so more people are more willing to purchase Christmas trees for celebration.

一張含有 文字, 行, 螢幕擷取畫面, 圖表 的圖片

自動產生的描述(4)

一張含有 文字, 螢幕擷取畫面, 行, 繪圖 的圖片

自動產生的描述

The graph on the left shows the correlation between the number of trees sold and the sales revenue, while the right one demonstrates the correlation between the average tree price and the sales revenue. By calculation in R, the former correlation equals 0.4947227, while the latter is 0.3126897. It implies that the correlation between the number of trees sold and sales revenue is more robust than price. Hence, if sellers want to generate higher profits, they should improve the quantity sold through promotion and advertising instead of direct markup.

Data Source:

<https://www.kaggle.com/datasets/thedevastator/us-christmas-tree-sales-data/data>