Awarding Body: **Arden**
Programme Name: **MSc Data Analytics and Information Systems Management**
Module Name (and Part if applicable): **Data handling and decision making**
Assessment Title: **Part 1: Essay**
Student Number: **STU105262**
Tutor Name: **Barry Smith**
Word Count: **1050** Please refer to the Word Count Policy on your Module Page for guidance

Summary

One of the most important components of developing workable scenarios for processing relevant data is **data handling and decision making**. This approach lays out a clear path for analysing and predicting future business and financial aspects. This interpretation aids in the proper formulation of sound judgments for the company's future ambitions. The right association of valid data analytics with the help of **Excel** has been provided in this study for our organization **Uber Technologies Inc.,** to overcome the supply-demand gap and a detailed display of graphs, charts, and reports was created using an associative procedural method to improve understanding of data manipulation for the required decision-making procedure. Below report shows that data assessment has an important role in making decisions for business development. By using the dataset from Kaggle we analyse the root causes for demand and supply gaps for rides and recommend realistic applicable solutions.

1.1 Organization: Uber Technologies Inc.,

Uber is an American multinational company providing cab hailing services across 900 cities worldwide (Uber - Wikipedia, 2021).A user can use a mobile app (Available for both Android and IOS users) or website and book a ride from destination A to B. Once the customer books a ride the app matches riders with available drivers(cars) and confirms the ride based on their availability. The main service provided by **Uber** is the ***technology*** where it provides a user application which connects passengers with drivers and facilitates automatic payments

1.2 Inspection of data integrity and potential gaps in data analytics and data protection.

Uber deals with big data generated by millions of people across the world which includes passenger personal details, location, destination, transactional details etc, that need to be stored with provable completeness, compliance, and consistency. To guarantee the data integrity uber uses **ledgerStore (Patel, Chovatia and Devarajaiah, 2021)**

Data breaches can be avoided by using **zero-trust approach** that only grants access to services based on what we know about the user and their device. A zero-trust stance that ensures all access to services must be authenticated, authorized, and encrypted. (Williams, 2021)

For data protection GDPR policies and guidelines should be followed to avoid any legal obligations.

The present challenge uber is facing is with the cancellation of rides from airport to city or vice versa either by rider or passenger because of cost factors or excess waiting time. To address the problems in supply-demand across various hours for these routes, a data gap analysis has been performed to make the best use of the organizational data which can benefit to overcome demand-supply issues.

Formatted: Font color: Auto, Not Expanded by / Condensed by , Pattern: Clear

Table 1.1 GAP ANALYSIS TO OVERCOME SUPPLY-DEMAND ISSUE.

Problem	Demand-Supply from airport to city routes.
Goals	Minimise cancellation to meet the demand supply.
Gaps (Causes)	Cancelled cars and availability of cars.
Solutions	Perform Exploratory data analysis and identify the gaps and provide actionable solutions.

Along with this SWOT analysis has been performed to identify critical success factors which should be maintained to keep up the market presence.

Table 1.2 Shows the SWOT analysis for uber on basis of its strengths, weaknesses and areas of scope and improvement.

Table 1.2 SWOT ANALYSIS

Strengths	**Weaknesses**
• Low operational costs (Interactions are only from driver to passenger) • Convenient and low cost compared to traditional taxis.	• Inconsistent response times and surcharges leading to ride cancellation effecting revenue. • Data Privacy & integrity Issues. (Uber records where the customer gets the taxi from and where they go) • Totally Relied on Technology.

Opportunities	Threats
• Introducing new Transport options (E-bikes, Motorcycles, peer-to-peer car rentals and autonomous driving) • Implement ML and AI techniques to change things across Transportation & mobility to customer support and driver partner navigation.	• Different legal regulations in different countries might affect the operations. • Self-Driving cars like Google's Waymo project might reduce the need for Uber.

1.3 Key data sources & Data sets available:

Uber uses both **internal** and **external** data for it its operations to predict the charges automatically using Traffic and GPS data by using machine learning algorithms which will predict the estimated journey time before the trip.

Data from its own network generated by their own apps from driver and rider is also an integral resource. **External data** like transportation routes can also be used to plan various services.

Uber obtains the necessary information from two sources: the application (client) and the backend services used by the app. Client logs are either created automatically by the platform (e.g., user interactions with UI components, impressions, etc.) or explicitly added by developers. More metadata is available in backend logs, which is either unavailable for mobile or too much for a mobile phone to manage. Each logged record has a key that allows it to be linked to the mobile interaction, resulting in a unified display. (Ravichandran and Verma, 2021)

2.1 Reorganisation of the current data-driven processes to streamline and enhance the data analytics and decision making.

To answer the **supply-demand gaps** to overcome the surge prices, cancellation, and high waiting time uber needs to implement **machine learning algorithms** (Richman, 2021) to predict the high demand areas by using **heat maps** (Shou, Chen and Zheng, 2021) then drivers can wait in that location which can reduce the surge charges, cancellation, and high waiting times. Data driven process can further improvised by using telematics (**Advanced fleet tracking**) (Sotra, 2021) where data related to GPS location, accelerometer, gyroscopic, engine diagnostics, idle time, fuel consumption, harsh breaking etc and integrate it with the mobile application and software systems and analyse it to minimise the supply-demand gaps. Data generated can be used to check if the driver is working with any other competitor, optimisation of routes, driver, and passenger safety, idling (to reduce carbon footprint) all these factors can help to fill the gaps to achieve the goal of demand-supply.

By introducing hyperscale, real-time matching model which is combination of digital platforms with location-based mapping technology. Passengers are matched to the nearest driver with highspeed, and location data can be used to check the fluctuations in supply and demand. Allowing dynamic pricing adjustments, for drivers by creating incentives to work more during peak hours. Which will encourage more drivers to join and work during the high demand hours unleashing flexible supply. (kinsey, 2021)

Hence, by analysing results from these methods decision-making can help to streamline internal business process, identify consumer trends, and mitigate the risks involved thereby improving the transportation overall.

2.2 Ubers Road map for next generation of big data platform.

To enhance big data platform for improved -data quality, data latency, efficiency, scalability, and reliability below are the steps to be followed: (Shiftehfar, 2021)

Data Quality: To prevent an influx of bad data, Uber should perform schema checks on data content and reject data entries if there is any identified issue. (e.g., not confirming with the schema)

Data Efficiency: To improve data efficiency Uber need to move away from relying on dedicated hardware for any of services towards service dockerization. Adoption of cloud-native applications to build applications that scale well on cloud infrastructure and operatable across bigdata with lower operational costs.

2.3 Compliance aspects of the proposed changes in data analytics.

Uber business model uses Machine learning techniques to analyse the user data and improve user experience by making use of continuous feedback from its users. As per EU General Data Protection Regulation (GDPR) collection of user information should be limited to an extent in such a way that data useful to fulfil a certain purpose should be collected and analysed. But in complex machine learning models it is difficult to decide the minimal amount of data required. This can be achieved by removing some of the input features. By using the knowledge within the model to produce a generalized output which has minimal impact on its accuracy thereby data minimization can be achieved in a provable manner. (Goldsteen et al., 2021)

3.1 Proposed Big data analytics:

- Use of Machine learning models to predict the high demand areas by using heat maps to point the high demand locations.
- By using vehicle telematics different data ranging from GPS location to detailed vehicle information can be gathered and transferred to the software systems which can be analysed in detail.
- Implementing hyperscale real-time data driven models.

3.2 Range of business decision that can be supported by proposed data analytics:

By using **ML and hyperscale real time** matching predictive techniques driver can be notified to place himself in the areas of high demand and based on the location-based mapping technology drivers can be mapped to nearest drivers within seconds. Reducing the waiting time and surcharges which will eventually reduce the number of cancellations. From vehicle telematics data driver performance can be continuously monitored and analysed. Then we can cut off the poor performing drivers by considering factors like harsh breaking, over speeding etc ensuring the driver and passenger safety.

3.3 Decision formulation for meeting supply-demand.

The major issue Uber is dealing is to meet the supply-demand from airport to cities and vice versa which can be done by using different analytics techniques by using heat maps to know **high demand area** and **time(Morning, afternoon, evening, night)** at which Uber drivers are receiving more requests. Drivers and employees play a major role to solve this issue. Conduct driver orientation sessions and inform them of the insights gained from the problem's analysis so that they know where to be and when to be there.

References

Goldsteen, A., Ezov, G., Shmelkin, R., Moffie, M. and Farkash, A., 2021. *Data Minimization for GDPR Compliance in Machine Learning Models*. [online] *arXiv.org. Available at: <https://arxiv.org*/abs/2008.04113> [Accessed 18 November 2021].

kinsey, m., 2021. *Data-driven cities ? - ibicity*. [online] Ibicity.fr. Available at: <http://www.ibicity.fr/data-driven-cities/?print=print> [Accessed 17 November 2021].

Patel, P., Chovatia, J. and Devarajaiah, K., 2021. *How Uber Migrated Financial Data from DynamoDB to Docstore*. [online] Uber Engineering Blog. Available at: <https://eng.uber.com/dynamodb-to-docstore-migration/> [Accessed 16 November 2021]

Ravichandran, D. and Verma, V., 2021. *How Data Shapes the Uber Rider App*. [online] Uber Engineering Blog. Available at: <https://eng.uber.com/how-data-shapes-the-uber-rider-app/> [Accessed 16 November 2021].

Richman, D., 2021. Q&A: Uber's machine learning chief says pattern-finding computing fuels ride-hailing giant. [online] GeekWire. Available at: <https://www.geekwire.com/2016/uber-collapse-without-pattern-finding-computers-says-chief-machine-learning/> [Accessed 17 November 2021].

Shou, X., Chen, P. and Zheng, H., 2021. *Predicting the Heat Map of Street Vendors from Pedestrian Flow through Machine Learning*. [online] Available at: <https://www.researchgate.net/publication/350447175_Predicting_the_Heat_Map_of_Street_Vendors_from_Pedestrian_Flow_through_Machine_Learning> [Accessed 17 November 2021].

Sotra, M., 2021. *Taking telematics to the next level: advanced fleet tracking | Geotab*. [online] Geotab. Available at: <https://www.geotab.com/blog/advanced-fleet-tracking/> [Accessed 17 November 2021].

Shiftehfar, R., 2021. *Uber's Big Data Platform: 100+ Petabytes with Minute Latency*. [online] Uber Engineering Blog. Available at:

<https://eng.uber.com/uber-big-data-platform/> [Accessed 9 December 2021].

Shao, Z. and Islam, M., 2021. Challenges and Opportunities to Dramatically Reduce the Cost of Uber's Big Data. [online] Uber Engineering Blog. Available at: <https://eng.uber.com/challenges-opportunities-to-reduce-cost-ubers-big-data/> [Accessed 17 November 2021].

Williams, C., 2021. [online] Centrify.com. Available at: <https://www.centrify.com/blog/the-uber-data-breach/> [Accessed 16 November 2021].