MGT7178: Data Management

ASSIGNMENT 2

Name: Dhanush Mathighatta Shobhan Babu
Student ID: 40412492
Word Count:2173

Lyft: A ride whenever you need one

Title : Driving Forward: An In-Depth Analysis of Lyft's Big Data Application, Technical Infrastructure, and Market Dynamics

Table Of Content

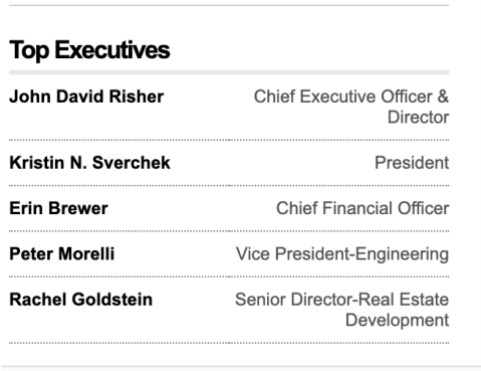
1. Introduction

Company Overview:

Lyft, launched in 2007 as Zimride, is a major transportation-as-a-service company. It operates mostly in the US and Canada (Yahoo Finance, 2023). The company offers a flexible platform for people to book last-minute travels. This platform offers many travel options through a multimodal network (GlobalData, 2023). The Ridesharing Marketplace connects drivers and customers, Express Drive offers flexible car rentals, Lyft Rentals handles long-distance rides, and a network of shared bikes and scooters for short trips is part of this network.

The software adds third-party public transit data, boosting riders' mobility options (Yahoo Finance, 2023). Lyft also offers autonomous vehicles, enterprise transportation solutions, Lyft Pink, and Lyft Pass commuter programmes. Marcus Cohn, John Zimmer, Rajat Suri, Matt van Horn, and Logan Green founded the social rideshare platform, which emphasises community connection (Forbes, 2023).

Lyft's CEO and Director is John David Risher, President is Kristin N. Sverchek, CFO is Erin Brewer, VP-Engineering is Peter Morelli, and Senior Director-Real Estate Development is Rachel Goldstein (CNN Money, 2023). Corporate ownership is split between institutional (42.41%), mutual fund (38.08%), and individual (28.03%) stakeholders (CNN Money, 2023). Lyft manages its many services and platform with 4,675 employees (Forbes, 2023).



Top Executives

John David Risher	Chief Executive Officer & Director
Kristin N. Sverchek	President
Erin Brewer	Chief Financial Officer
Peter Morelli	Vice President-Engineering
Rachel Goldstein	Senior Director-Real Estate Development

Fig 1: Pictorial representation of Top executives in LYFT
Source : https://money.cnn.com/quote/profile/profile.html?symb=LYFT

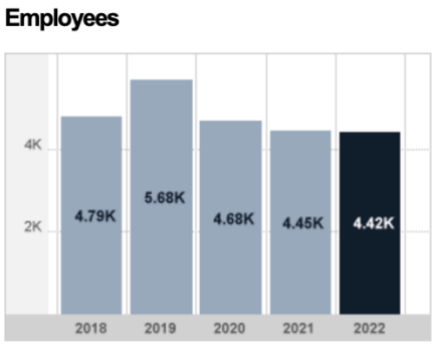


Fig 2: Pictorial representation of Employee in LYFT year-wise
Source : https://money.cnn.com/quote/profile/profile.html?symb=LYFT

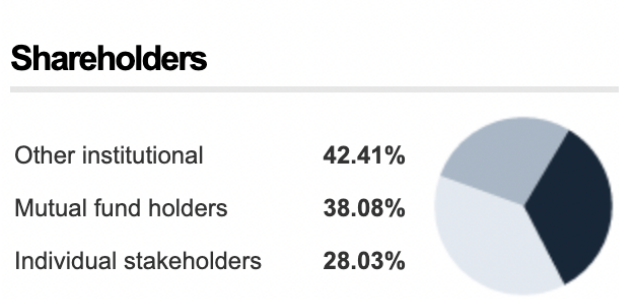


Fig 3: Pictorial representation of Shareholders in LYFT
Source : https://money.cnn.com/quote/profile/profile.html?symb=LYFT

Market analysis of Company:

As of 2022, Lyft Inc. has a $18.72 billion market valuation and $3.62 billion revenue. From 2018 to 2020, Lyft's revenue grew by $2 billion (DC Finance, 2023).

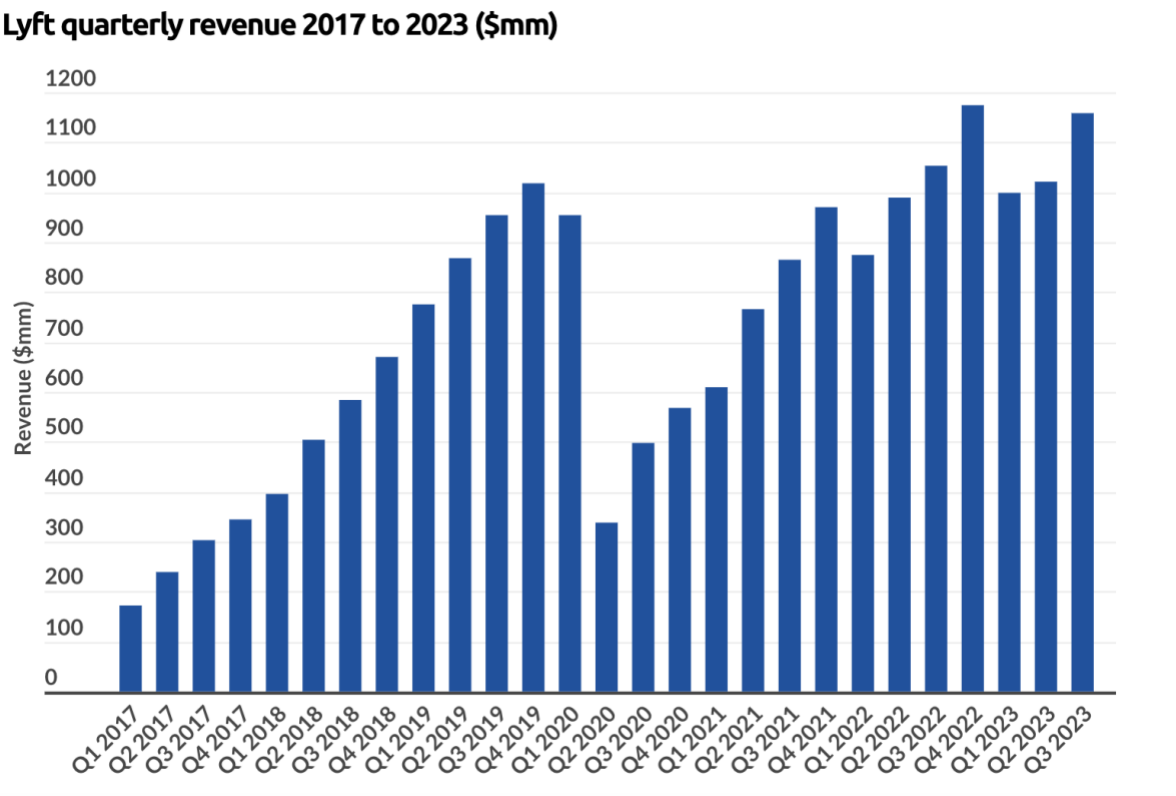


Fig 4: Graphical representation quarterly revenue of LYFT
Source : https://www.businessofapps.com/data/lyft-statistics

Lyft annual revenue 2017 to 2022 ($bn)

Year	Revenue ($bn)
2017	1.05
2018	2.15
2019	3.61
2020	2.36
2021	3.2
2022	4.09

Fig 5: Tabular representation quarterly revenue of LYFT
Source : https://www.businessofapps.com/data/lyft-statistics/

Lyft's net losses increased by $522 million in 2022, continuing its streak of losses.

Lyft annual net profit/loss 2017 to 2022 ($mm)

Year	Net income/loss ($mm)
2017	-688
2018	-911
2019	-2602
2020	-1753
2021	-1062
2022	-1584

Fig 6: Tabular representation of LYFT Profit/Loss
Source : https://www.businessofapps.com/data/lyft-statistics/

Lyft's average revenue per active rider climbed 11.4% to $57.72 in Q4 2022.

Lyft active rider ARPU 2017 to 2022

Year	APRU
2018	36.02
2019	44.40
2020	45.40
2021	51.79
2022	57.72

Fig 7: Tabular representation of LYFT ARPU
Source : https://www.businessofapps.com/data/lyft-statistics/

Lyft, Inc. (Lyft) delivers food and transportation by smartphone (Appiah, 2022). It runs in 645 US communities and 10 Canadian locations. They are more important in Canada than the US (Appiah, 2022).

In Q4 2022, Lyft recorded 20.3 million active passengers; this number appears to be approaching stability, as it remained unchanged in Q3 2022.

Lyft quarterly users 2016 to 2023 (mm)

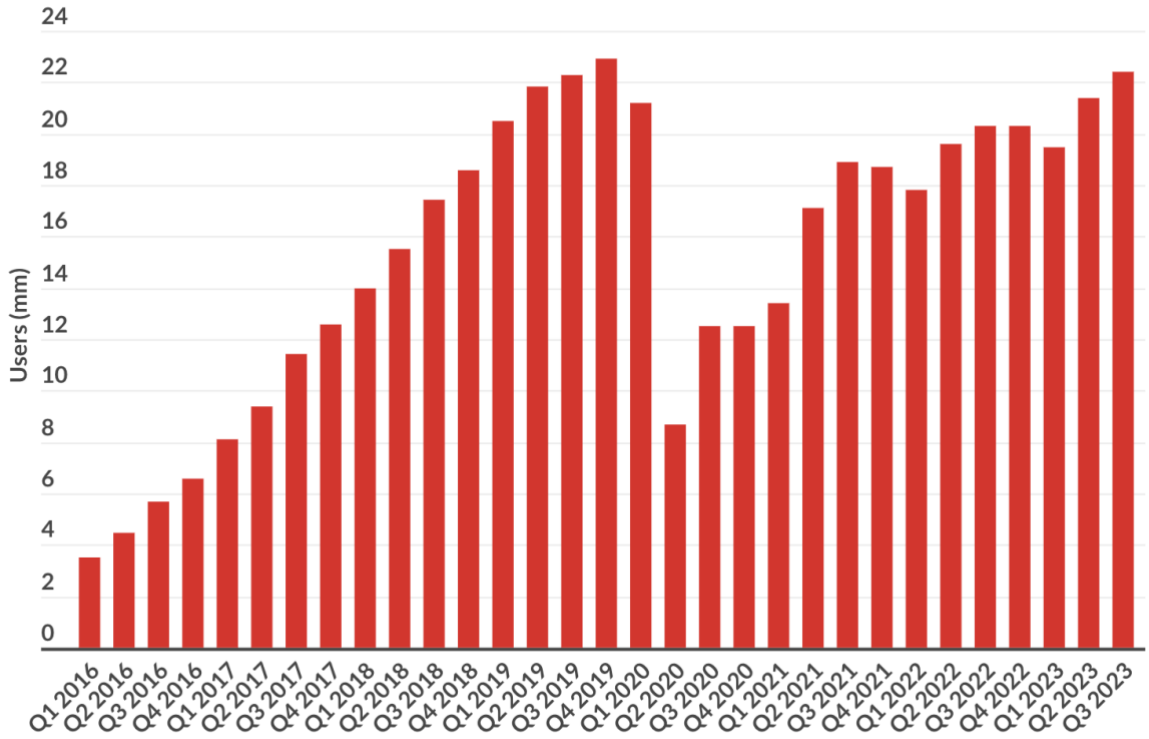


Fig 8: Graphical representation of LYFT Active users
Source : https://www.businessofapps.com/data/lyft-statistics/

Lyft annual users 2016 to 2022 (mm)

Date	Users (mm)
2016	6.6
2017	12.6
2018	18.6
2019	22.9
2020	12.5
2021	18.7
2022	20.3

Fig 9: Tabular representation of LYFT Active users
Source : https://www.businessofapps.com/data/lyft-statistics/

Financial metrics suggest a $6.11 billion market valuation for the corporation. Value indicators include a 34.30 forward P/E ratio, 1.35 price-to-sales ratio, and 13.05 price-to-book value ratio (Forbes, n.d.).

Fig 10: Representation of LYFT Stock price
Source : https://www.businessofapps.com/data/lyft-statistics/

Lyft has about 29% market share and a large network. It is the second largest US corporation after Uber (IQBAL, 2023).

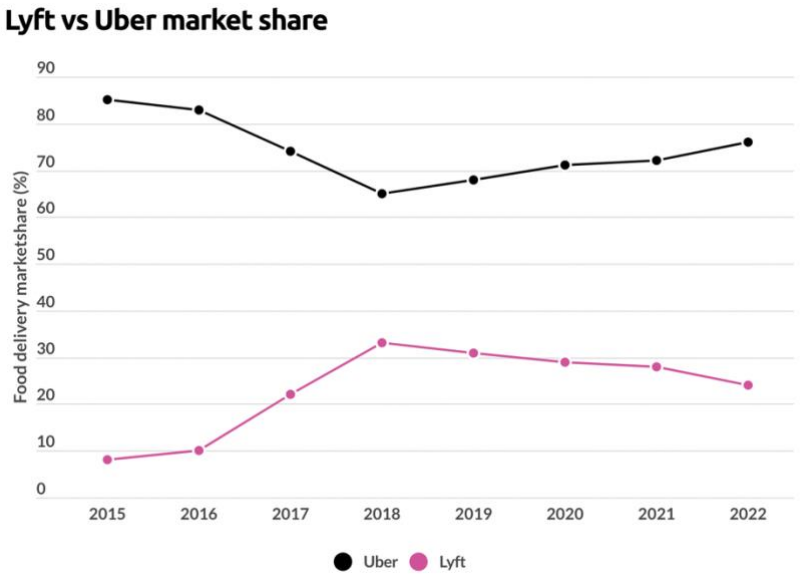


Fig 11: Uber vs LYFT Market Share
Source : https://www.businessofapps.com/data/lyft-statistics/

2. Big Data Application Overview in LYFT

According to (Mirchandani, 2020) Big Data Applications at Lyft are:

1. **Marketplace**: Optimizing ride dispatch and pricing through data-driven models.
2. **Rider**: Enhancing rider experiences through data analysis and personalized programs.
3. **Driver**: Developing strategies for driver retention and engagement via data-driven insights.
4. **Rideshare Planning & Operations**: Using data tools for business diagnostics and forecasting.
5. **Mapping**: Improving accuracy and quality of Lyft's map through data science.
6. **Customer Platforms**: Applying data models for payments, support, fraud, and insurance.
7. **Lyft Business**: Utilizing frameworks to enhance product experiences and partnerships.
8. **Research (Marketplace Labs & Economics)**: Solving complex problems using scientific expertise.
9. **Fleet**: Optimizing growth and profitability through data-driven pricing and forecasting.
10. **TBS (Transit, Bikes & Scooters)**: Enhancing operational efficiency using data analysis.
11. **Level 5 (Autonomous)**: Utilizing data models to advance self-driving vehicle deployment.
12. **Recommendation System:** Implementing data-driven recommendation algorithms to personalize ride and service suggestions for Lyft users (Niu, 2023).

Our focus will be on the **recommendation system**.

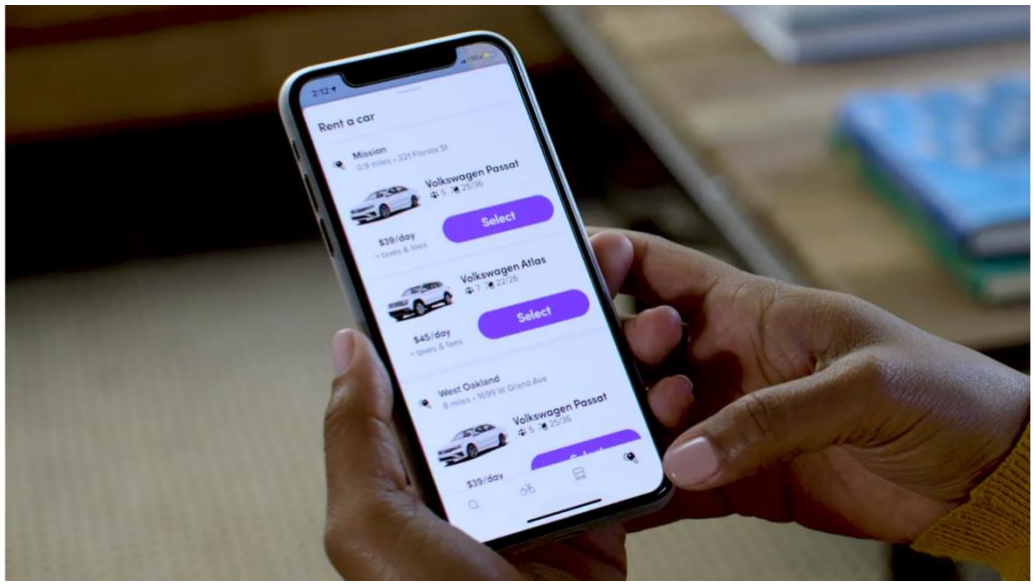


Fig 12: Cab recommendation in LYFT app
Source : https://eng.lyft.com/what-is-data-science-at-lyft-4101a69be028

Recommendation system in LYFT

Recommendation is crucial for Lyft to comprehend its riders and enables the customisation of app experiences to better meet their requirements. Occasionally, recommendations are utilised to regulate the marketplace, ensuring a harmonious equilibrium between the demand for rides and the availability of drivers (Niu, 2023). This enables ride requests to be satisfied with more favourable dispatch outcomes, such as pairing riders with the most suitable driver in proximity (Niu, 2023).

Scope and function of the Recommendation System

1**. One-Tap Module:** Designed for Lyft users who frequently travel and have a significant ride history, this feature allows them to rapidly book a ride with just one tap on the home page. By bypassing the usual request routine, this feature saves time for users who are in a rush (Niu, 2023).

2**. Ranking & Preselection:** Once the location is determined, customers are provided with a prioritised assortment of transportation choices, which are determined by their personal preferences and the prevailing market conditions. Visual cues, such as the "fastest" and "your usual" transportation options, assist consumers in making expedited choices (Niu, 2023).

3.Lyft provides **post-request upgrade** choices to riders in some situations, such as improved estimated time of arrival or rates, through a popup that appears after the request is made (Niu, 2023).

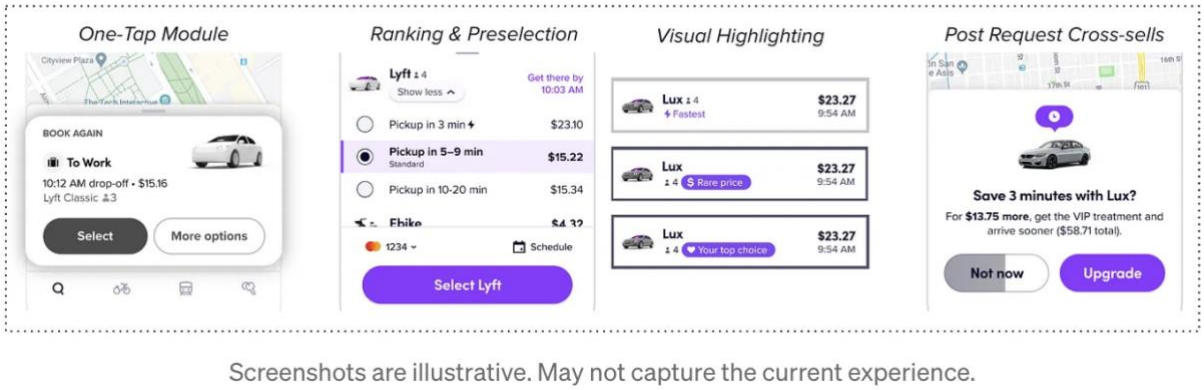


Fig 13: Recommendation in LYFT app
Source: https://eng.lyft.com/the-recommendation-system-at-lyft-67bc9dcc1793

Evolution of Mode Recommendations:

Challenges addressed by (Niu, 2023) are as follows :

1. **Over choice:** With a growing portfolio, users faced difficulty in discovering and choosing from multiple modes

2. **Cold Start Problem:** New modes lacked visibility due to limited user interaction.

3. **Dynamic Goals:** Balancing user experience, marketplace equilibrium, and ride reliability beyond mere rides and conversions.

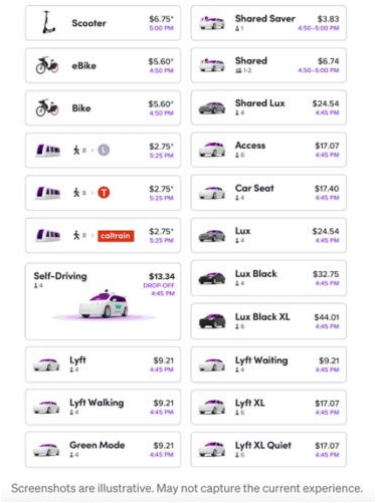


Fig 14: Challenge of Overchoice
Source: https://eng.lyft.com/the-recommendation-system-at-lyft-67bc9dcc1793

Solution Overview used by LYFT (Niu, 2023) :

1. **Over choice Resolution**: Leveraging machine learning models to predict user propensity for different modes, customizing rankings based on rich information like location, time, ride history, and preferences.
   - **Algorithm**: LightGBM (Machado et al., 2019), employing lambda rank or multi-class classification based on use cases.
   - **Hyperparameters**: Extensive tuning using Lyft's distributed hyperparameter optimization pipeline.

2. **Preselection & Cognitive Load Reduction:** Initially limiting mode display to 3–4 options above the fold to streamline user focus, while enabling exploration below.
   - Preselection Evolution: Moved from last mode taken to most frequently used mode to a model-based approach predicting highest propensity.

3. **Cold Start Problem Mitigation:** Introduced a post-processor to adjust model results for new offerings with limited user data, enhancing their visibility.

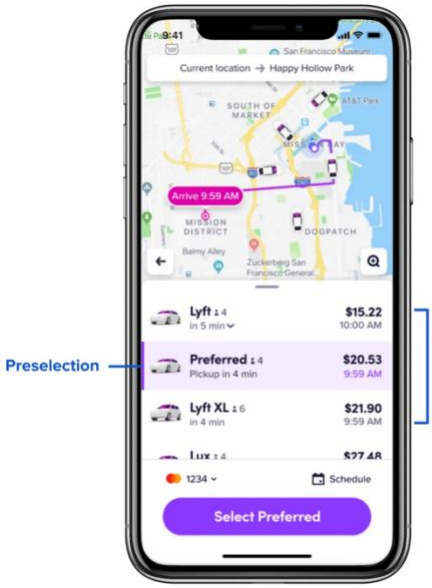


Fig 15: Selection Cab
Source: https://eng.lyft.com/the-recommendation-system-at-lyft-

4.**Dynamic Goals Integration:** Optimising beyond rides and conversions by considering metrics like ride reliability and supply/demand balance.
   - Cross-Sell Experience: Rare but strategic post-ride prompts to balance demand and enhance matching efficiency, considering user experience factors.

Types of Data Utilized and Data warehousing in LYFT:

According to McPhillips (2023), tech companies routinely process petabytes of data. A concise breakdown of the types of data Lyft used (Shrivastava, 2020):

1. **Structured Data (MySQL):** Handles transactional details, user information, CRM data, HR management, and payroll records.
2. **Unstructured Data (Amazon DynamoDB, Amazon RedShift, etc.):** Stores GPS ride tracking, customer insights, real-time bookings, and event funnelling information.
3. **External Data Sources (APIs - Google Maps, Salesforce, etc.):** Integrates external APIs for location, routes, trip cost estimation, diverse ride options.
4. **Sensor Data (GPS, Traffic, Weather):** Collects sensor-driven information for GPS locations, traffic conditions, and weather attributes to enhance ride experiences and decision-making.
5. Lyft has **semi-structured** data on complicated driver-rider-location-time interactions (Xu, 2023).

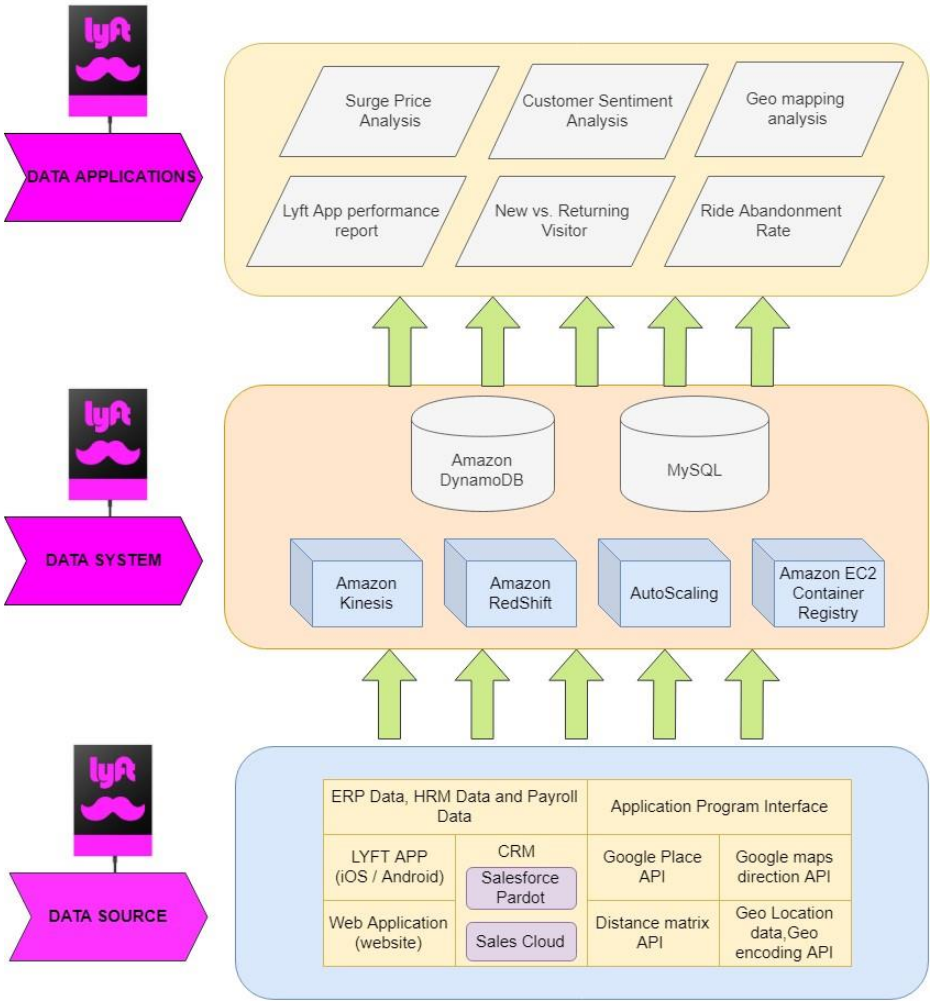
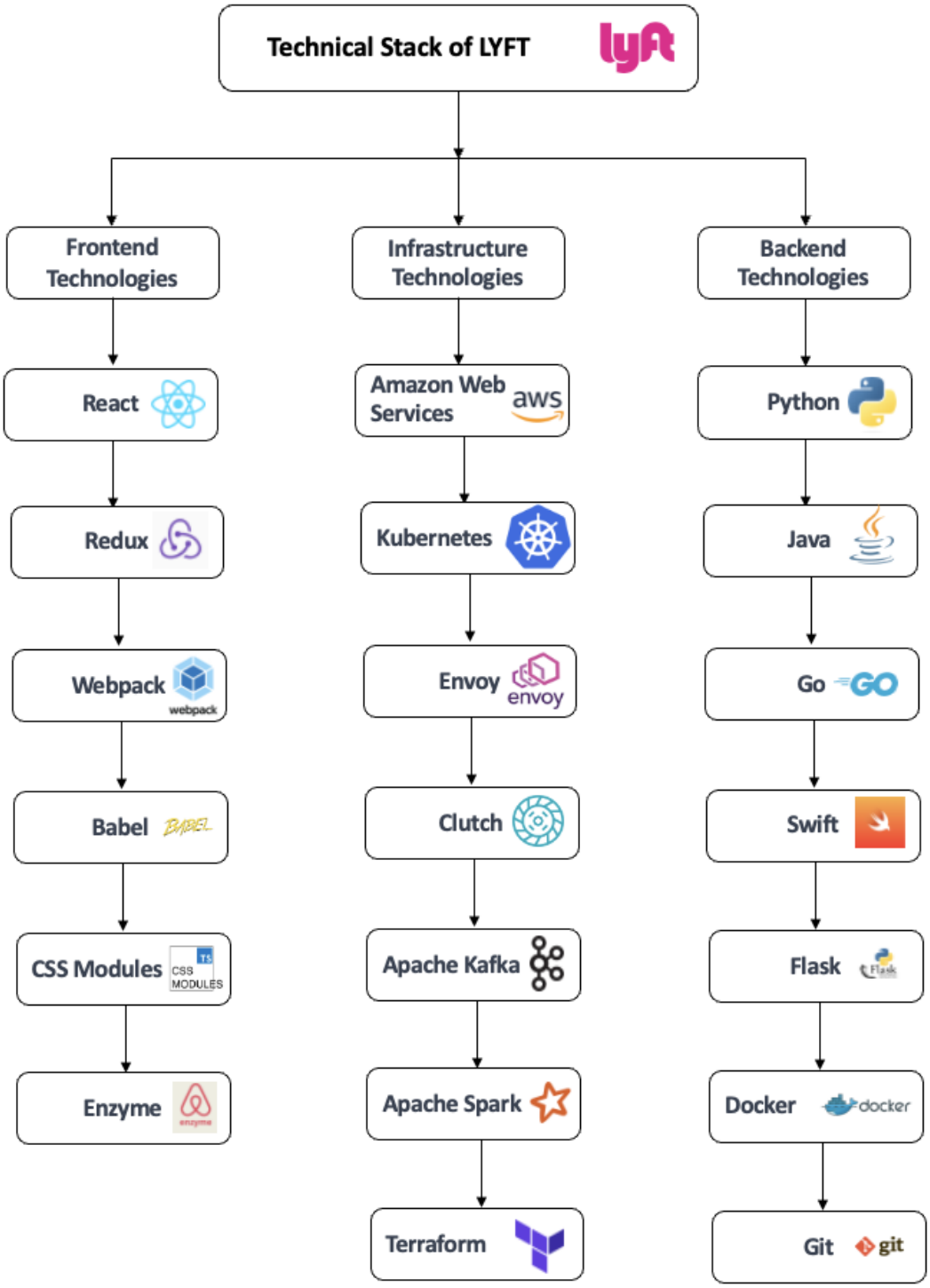


Fig 16: Data Warehousing
Source: https://www.skillsire.com/read-blog/241_data-warehousing-in-lyft.html

3. Technical infrastructure used in LYFT:

According to Arjun (2023), the technologies employed by LYFT can be visualised in the following flow chart:



Data Quality in LYFT:

McPhillips (2023) emphasizes the critical role of high-quality data in data-driven companies and introduces Verity as Lyft's proprietary solution for ensuring data quality.

Data quality includes semantic correctness, consistency, completeness, well-formedness, and timeliness to assure accurate, reliable data for intended purpose. It ensures reliable decision-making, operational efficiency, and reliable analytics, decreasing errors and improving corporate operations (McPhillips,2023).

Fig 17: What is Data quality?
Source: https://eng.lyft.com/from-big-data-to-better-data-ensuring-data-quality-with-verity-a996b49343f6

The Hive Data Warehouse is Lyft's main data analysis store. It stores data from millions of daily events. It is essential for analytical activities like ETL and model training, which drive machine learning models and predict market trends (McPhillips,2023). Although important, it had data quality issues that affected metrics and efficiency, underlining the need for good data quality management.

Analytical Tasks in Lyft's Data Lifecycle:

Kafka and Flink process vast amounts of event data every day before storing it in Hive for analysis. This lifecycle helps ETL activities, which generate key datasets for model training, experimentation, and trend analysis. It underpins Lyft's business insights and operations (McPhillips,2023).

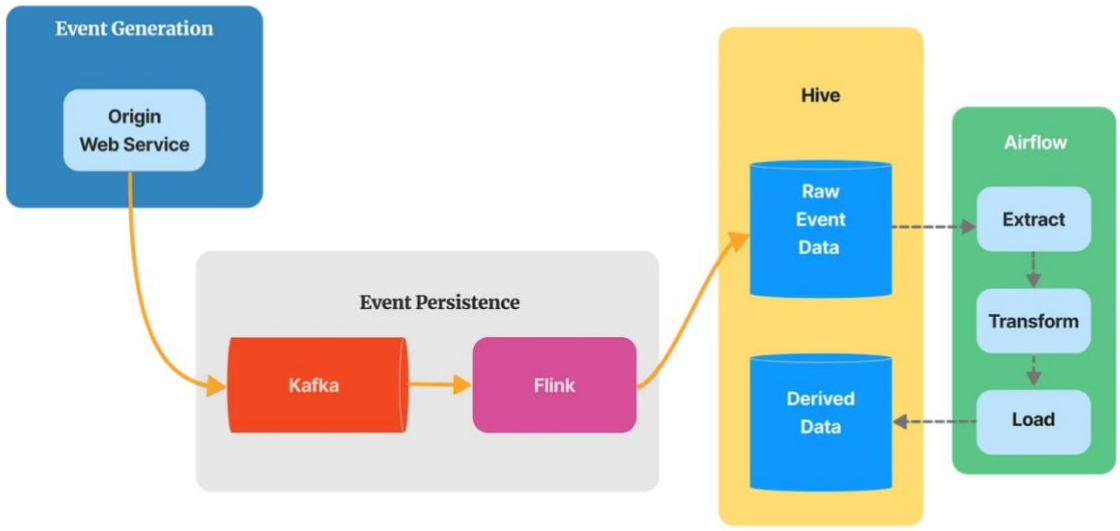


Fig 18: Analytic Event Life cycle is LYFT?
Source: https://eng.lyft.com/from-big-data-to-better-data-ensuring-data-quality-with-verity-a996b49343f6

Lyft's Hive Data Warehouse, crucial to its analytical processes, has incomplete information, semantic problems, and processing delays. These flaws caused skewed measurements, erroneous machine learning, and operational inefficiencies (McPhillips,2023). Verity, their in-house data quality platform, centralised data quality assessment to address these issues (McPhillips,2023). Verity checked data integrity for completeness, accuracy, and timeliness. This approach increased data reliability, reduced the impact of faulty data on analytical processes, and improved machine learning accuracy and Hive Data Warehouse operational efficiency.

Lyft's data quality platform, Verity, implemented a strong data quality verification and execution mechanism. It used user defined YAML Check Definitions to describe queries,

conditions, metadata, and notifiers (McPhillips,2023). These definitions enabled interactive testing and validation before deployment, assuring robustness.

Fig 19: User story of Verity
Source: https://eng.lyft.com/from-big-data-to-better-data-ensuring-data-quality-with-verity-a996b49343f6

Verity's Scheduler, API Server, Executor, and Notifier organised asynchronous compute engine execution. The Scheduler dispatched Verity Scheduled Checks independently from data orchestration engines to maintain performance (McPhillips,2023). The API Server used DynamoDB to efficiently store and retrieve external API checks and results. The Executor used Simple Queue Service (SQS) for scalability and fault tolerance to execute TrinoSQL, validate conditions, and store results (McPhillips,2023). For disruption-free notifications, the Notifier service separately retried unsuccessful alerts.

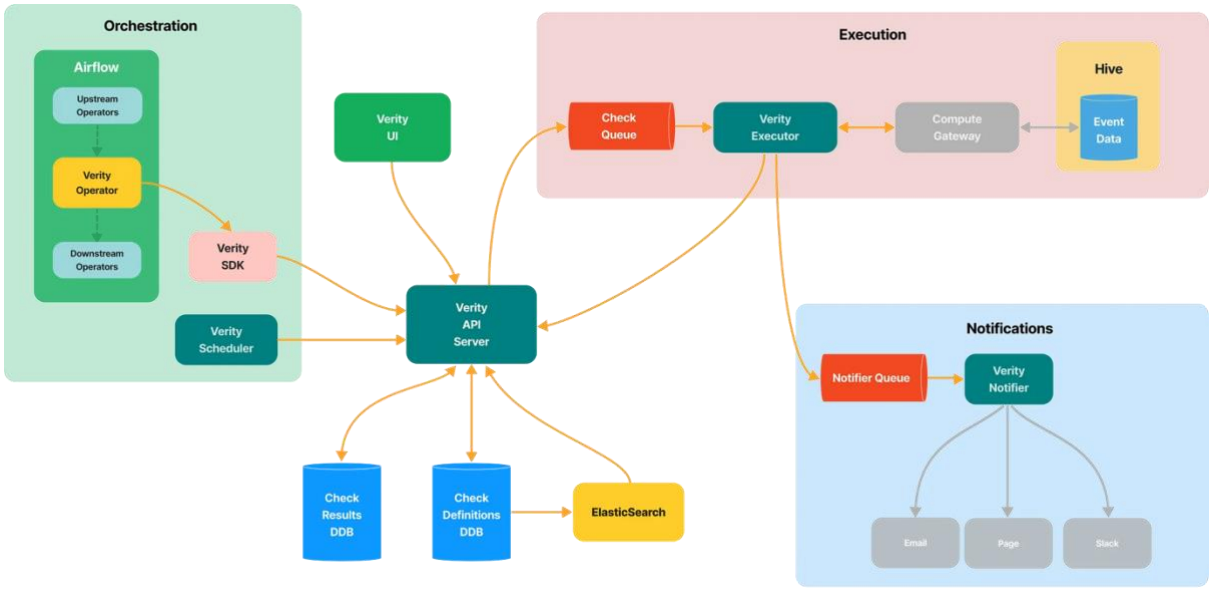


Fig 20: Execution of Verity
Source: https://eng.lyft.com/from-big-data-to-better-data-ensuring-data-quality-with-verity-a996b49343f6

This well-structured execution methodology helped Verity assess Lyft's data ecosystem for quality, dependability, and scalability.

LyftLearn: ML Model Training Infrastructure built on Kubernetes

The Lyft app relies on ML for dispatch, pricing, fraud detection, support, and more. Each team uses a different modelling technique, but a common platform is needed to simplify model development, parallelize model training, track past training runs, visualise performance, run models on schedule for retraining, and deploy trained models for serving.

Kakade (2021) has listed Key Features as follows :

1. **Fast Iterations:** Evaluating and iterating ML issue approaches quickly.
2. **Library Flexibility:** Except for security, no limits on ML libraries or versions.
3. **Layered Cake Approach:** Offering API, CLI, and GUI access for user preferences.
4. **Cost Visibility:** Helping users analyse business effect by showing training run costs.
5. **Usability:** Making distributed training and hyperparameter tuning self-serve and integrated with existing solutions.

Architectural Parts (Kakade,2021):

1. **Model Development:** Users can use Jupyter, R-studio, or their preferred editor. They create development environments with hardware setups and base images. Version-specific models can be preserved as containers.
2. **Training and Batch Prediction:** API or CLI/GUI jobs can be configured. Kubernetes parallelizes training jobs for diverse configurations, enabling retraining.
3. **User Dashboard:** Displays models, versions, training runs, performance data, logs, and production deployment.
4. **Image Build:** Users can extend base pictures to build team-specific images.
5. **Data and Compute:** Kubernetes, AWS Elastic File System, Lyft's data warehouse, and AWS RDS Aurora metadata.

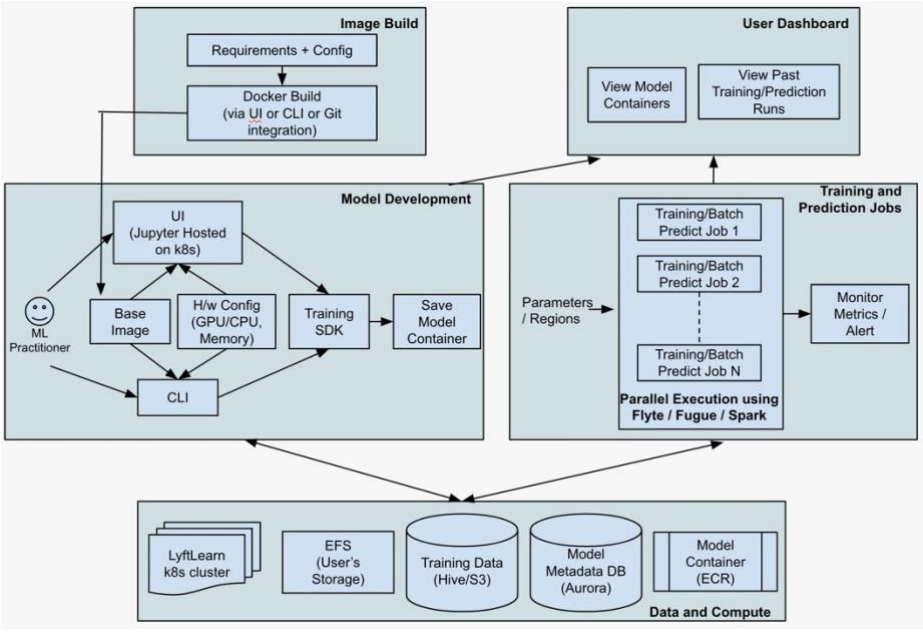


Fig 21: Architectural layout of Lyft learn.
Source: https://eng.lyft.com/lyftlearn-ml-model-training-infrastructure-built-on-kubernetes-aef8218842bb

In summary, LyftLearn's architecture supports many ML applications within Lyft, enabling quick model construction, training, and deployment. Flexibility, speed, and interaction with current infrastructure make it a significant component of the company's ML pipeline across teams.

4. Benefits to LYFT and competitive advantage

Lyft possesses a robust information system that efficiently handles, retains, and disseminates the information required for effective decision-making. Lyft possesses a robust information system that efficiently handles, retains, and disseminates essential information required for effective decision-making.

Software, hardware, and telecommunications networks are integrated and harmonised to enable system functions and operations (Bourgeois, 2014; Madakam et al., 2015).

Lyft utilises Amazon Web Services and Cloud technology to effectively handle fluctuations in business demand (Appiah, 2022). Lyft utilises data science technologies to effectively leverage historical data for machine learning and strategic planning purposes (Pialat, 2020). This enhances Lyft's capacity to provide competitive pricing and guarantee an exceptional client experience. There is extensive synergy between the service and development teams, including the engineering and analytics departments of the organisation (Appiah, 2022).

Cloud Computing: Amazon Web Services and substantial cloud technology investment help Lyft manage industry hurdles and business spikes (Madakam et al., 2015). Lyft relies on data science techniques like **AI and machine learning.** These tools help price, estimate driver/passenger availability, improve services, and satisfy customers (Appiah, 2022).

Open Autonomous Vehicle Platform: Lyft wants autonomous vehicles to replace drivers and improve safety (Mirchandani, 2020).

Lyft's success is also attributed to cultivating a culture that places importance on the genuine connection between drivers and riders, which has greatly contributed to its expansion and customer contentment (Appiah, 2022).

5. Limitations to LYFT

Generic Problems Identified by Appiah (2022) are:

- **Drivers without cars:** Around 20% of drivers don't own vehicles, impacting efficiency and earnings.
- **High gas prices:** Costs eat into driver profits, affecting their income.
- Delayed commission payments: Inconvenience caused to drivers due to slow payment processes.
- **Reports of sexual assault:** troubling cases reported, especially affecting female passengers(Bussewitz, 2020).
- **Lack of in-house delivery platform:** Unlike competitors, Lyft lacks its own mobile delivery ordering system.
- **Market concentration:** Overreliance on limited markets, primarily the USA and Canada.

Ethical Concerns by Ley (2018):

1. **Racial Bias:** Studies suggest racial disparities in Lyft wait times, challenging the fairness of service despite anti-discrimination policies.

2**. Driver Treatment:** Contractor classification raises concerns about benefits and protections, akin to issues seen in Uber.

3. **Transparency and Manipulation:** Reports hint at possible user manipulation tactics, raising doubts about Lyft's fairness and transparency.

4. **Ethical Obligation:** Lyft needs to tackle discrimination, driver welfare, and transparency issues for a more ethical service.

Some big data limitations are Lyft's high operating costs are due to the amount of data company processes (Saha, 2023). With rising data resources, it becomes harder to identify what they are, how to access them, and what information they include (Feng, 2019).

6. References

- Arjun (2023) Inside lyft tech stack and infrastructure, Appscrip Blog. Available at: https://appscrip.com/blog/lyft-tech-stack-and-infrastructure/#:~:text=management%2C%20and%20observability.-,Infrastructure%20technologies%20of%20Lyft%20Tech%20Stack,%2C%20and%20database%20management%2C%20respectively . (Accessed: 20 December 2023).

- Bourgeois, D.T. (2014) Information systems for business and beyond. Arlington, VA: Saylor Academy.

- Bussewitz, C. (2020) The anti-uber? lyft's track record on Safety, sexual assaults lawsuits undercuts 'good guy' image, USA Today. Available at: https://eu.usatoday.com/story/tech/2020/01/06/lyft-not-safer-than-rival-uber-sexual-assaults-drivers-critics/2828069001/ (Accessed: 20 December 2023).

- Feng, T. (2019) Open sourcing amundsen: A data discovery and metadata platform, Medium. Available at: https://eng.lyft.com/open-sourcing-amundsen-a-data-discovery-and-metadata-platform-2282bb436234 (Accessed: 20 December 2023).

- IQBAL, M. (2023) Lyft Revenue and usage statistics (2023), Business of Apps. Available at: https://www.businessofapps.com/data/lyft-statistics/ (Accessed: 17 December 2023).

- Kakade, V. (2021) Lyftlearn: ML Model Training Infrastructure built on Kubernetes, Medium. Available at: https://eng.lyft.com/lyftlearn-ml-model-training-infrastructure-built-on-kubernetes-aef8218842bb (Accessed: 20 December 2023).

- Ley, B. (2018) The usability, accessibility, and ethics of lyft, Medium. Available at: https://medium.com/@leybrandi/the-usability-accessibility-and-ethics-of-lyft-58599a2c2668 (Accessed: 20 December 2023).

- Lyft (2023) CNNMoney. Available at: https://money.cnn.com/quote/profile/profile.html?symb=LYFT (Accessed: 17 December 2023).

- Lyft | Lyft Stock Price, company Overview & News (2023) Forbes. Available at: https://www.forbes.com/companies/lyft/?sh=39898a86439a (Accessed: 17 December 2023).

- Lyft, Inc. (Lyft) Company Profile & Facts (2023) Yahoo! Finance. Available at: https://finance.yahoo.com/quote/LYFT/profile/ (Accessed: 17 December 2023).

- Machado, M.R., Karray, S. and de Sousa, I.T. (2019) 'LIGHTGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry', 2019 14th International Conference on Computer Science & Education (ICCSE)

[Preprint]. doi:Available at : https://ieeexplore.ieee.org/document/8845529 [Accessed on 19th December 2023].

- Madakam, S., Ramaswamy, R. and Tripathi, S. (2015) 'Internet of things (IOT): A literature review', Journal of Computer and Communications, 03(05), pp. 164–173. doi:Available at : https://www.scirp.org/journal/paperinformation?paperid=56616 [Accessed on 20th December 2023].

- Mirchandani, S. (2020) What is Data Science at lyft?, Medium. Available at: https://eng.lyft.com/what-is-data-science-at-lyft-4101a69be028 (Accessed: 18 December 2023).

- Niu, J. (2023) The recommendation system at lyft, Medium. Available at: https://eng.lyft.com/the-recommendation-system-at-lyft-67bc9dcc1793 (Accessed: 18 December 2023).

- Onasis Osei Appiah, I.. (2022) "TNCS' GROWTH STRATEGY: A CASE STUDY ON HOW LYFT CAN GROW ITS MARKET SHARE", Global journal of Business and Integral Security. Available at: https://www.gbis.ch/index.php/gbis/article/view/111 (Accessed: 17 December 2023).

- Pialat, G. (2020) How big data and IOT are connected, IoT For All. Available at: https://www.iotforall.com/big-data-iot (Accessed: 20 December 2023).

- Plc, G. (2023) Lyft Inc Company Profile - Overview, GlobalData. Available at: https://www.globaldata.com/company-profile/lyft-inc/ (Accessed: 17 December 2023).

- Ranger, S. (2020) What is the iot? everything you need to know about the internet of things right now, ZDNET. Available at: https://www.zdnet.com/article/what-is-the-internet-of-things-everything-you-need-to-know-about-the-iot-right-now/ (Accessed: 20 December 2023).

- Saha, A. (2023) Big savings on big data, Medium. Available at: https://eng.lyft.com/big-savings-on-big-data-9c74b7a35326 (Accessed: 20 December 2023).

- Shrivastava, A. (2020) Data Warehousing in lyft, Skillsire. Available at: https://www.skillsire.com/read-blog/241_data-warehousing-in-lyft.html (Accessed: 19 December 2023).

- The4 (2023) Marketing Mix Analysis of Lyft, Inc. (lyft)., dcf. Available at: https://dcf.fm/blogs/blog/lyft-marketing-mix (Accessed: 17 December 2023).