Flyber Data Strategy MVP

## **Introduction**

Flyber has been massively successful. Results have beaten expectations and projections! This is good news for Flyber, but now it’s time to plan for what's next. With success came some challenges. While we were able to grow, the original data pipelines to receive and process data are unable to keep up with the current and future growth.

As a Data Product Manager, working with multiple teams and stakeholders is imperative to success. To understand what our needs are, what scale we are growing at, and how we can build for the future, we need to consider all relevant stakeholders. In this proposal, present your findings along with the analysis and reasoning behind the choices made in order to help Flyber continue its success.

## **Section 1:** Data Customers & Needs

Flyber is a two-sided platform. You have customers who are riders, and you have partners who are drivers/pilots (think Uber: riders and drivers). For the Minimum Viable Product, you will be focusing on the Riders side of the business. To build an end to end data pipeline the very first step is to understand who needs data and why they need that data. Within Flyber, identify who your primary data customers/stakeholders are, why they are your primary data stakeholders and how they want to use the data (primary use-cases).

**Identify your primary internal stakeholders and their use-cases:**

*(You may add more rows if necessary.)*

|  |  |  |
| --- | --- | --- |
| **Stakeholder** | **Why are they primary stakeholders?** | **Use-Case** |
| Engineering | Designs and creates a product | * Monitoring website and app performance * Data generated by site and app * Alerts * Machine Learning and load prediction * Dashboard |
| Product Management | Plan the product life cycle, including the roadmap. Operates in the domain boundaries of engineering, business and human behavior. | * Identifying customer pain-points * Customer interaction data for site and app * Business Intelligence tool to analyse data * Big data use-cases * Tracking metrics to make product decisions * Visualization tools to tell stories * Understand your customers more effectively |
| Product Marketing | Identifies target customers and markets the product, including advertisement. | * Content creation * Targeted advertising * Virtual assistant * Customer profile and customer preferences * Just-sale, Business Intelligence, visualization and reporting |
| Sales | Sell the product. Need data to make sales forecasts monthly, quarterly and yearly. | * Sells and generate revenue * Reports for business reviews – monthly business reviews * Understand your customers more effectively * Drive performance and revenue * Identify sales trends * Provide personalized service more easily * Improve operational efficiency |
| Accounting | Responsible for keeping records of expenses and earnings. | * Reconciliation reports |
| Finance | Responsible for creating a financial outlook that will assist in making proper business decision | * Monitoring current Profit & Loss * Expense Report from Accounting, Business Intelligence and Visualization tool * Machine Learning for future finance predictions |
| Customer Service | Responsible for addressing customer needs as it relates to the product | * Provide personalized responses * Customer data across systems * Visual summary * Visualizations * Reports |
| Security & Risk | Protects the entity from fraudulent activities or ensures the integrity of business and customer information. | * Fraud Prevention |

## **Section 2:** Data Collection and Data Modelling

**To support our primary stakeholders’s use-cases we need the following data:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Stakeholder** | **Use-Case** | **Data** | **Why is this the primary use-case?** |
| Engineering | Monitoring website and app performance  Data generated by site and app | Event data  Event ID, Timestamp ,Event Type | Design and develop the app. Access to app performance data is crucial |
| Product Management | Identifying customer pain-points | Event data  Event ID, Timestamp, Event Type, Event Page | Need to understand the pain in order to create solutions to the problem. Also, planning and maintaining product life cycle |
| Product Marketing | Targeted advertising | Entity data  Customer name, Email, Phone, Address, Customer Order History | Need to reach the right audience at the right time using digital resources |
| Sales | Selling and generating revenue | Entity data  Customer name, Email, Phone, Address, Customer Order History | Must get the product to the consumer in order to generate revenue |
| Accounting | Keeping records of expenses and earnings. | Entity data  Number of Transactions, Cost of Transactions, Retail Price, Taxes, Earnings, others | Tracks income and expenditure for the business |
| Finance | Profit & Loss | Entity data  Aggregate Transactional Data, Number of Transactions, Cost of Transactions, Retail Price, Taxes | This is essential for business performance measure |
| Customer Service | Personalized response to the customer | Entity data  Order Number, Product Details, Cost, Price, Service Date | Being available to respond to customer needs ensure retention and extracts more value |
| Security & Risk | Prevent fraud | Event + Entity data | Ensure the integrity of the business and its objective |

**The tables we need are**:

*Note: As a best practice, we should establish these relationships between tables from the very beginning. To complete this exercise we will focus on fundamental concepts of relational databases - tables, normalization and unique keys. Please provide the table header row for each table, tables might be different lengths. Make sure you include the following for each table. You can create as many tables as you feel are necessary (copy and paste from one of the table sections):*

**[Customer ]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer\_ID [PK] | Name | Address | Zip | Email | Phone\_Number |

Rationale for Choosing Primary and Foreign Keys for the Table 1:

[The customer is at the center of the Flying Taxi business model. Customer information data is very crucial and necessary.

The relationship of the Customer\_ID in “Customer” table is one-to-many between “Order” table and “Payment” table. The Primary Key field cannot accept NULL values. The value is and will always be constant at all times]

**Table 2:**

**[Order]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Order\_ID [PK] | Customer\_ID[FK] | Order\_Date | Pickup\_Loc | Dropoff\_Loc | Employee\_ID |

Rationale for Choosing Primary and Foreign Keys for the Table 2:

[Following the use-case needs for transaction records, an “Order” table is required for booking Flying Taxi rides.

The Order\_ID uniquely identifies all stored order information in the “Order” table. For that reason, Primary Key for this table. The Foreign Key, which is Customer\_ID in this case is a Primary Key in the “Customer” table]

**Table 3:**

**[Payment]**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Payment\_ID [PK] | Customer\_ID [FK] | CreditCard\_Info | Billing\_Zip | Order\_ID [FK] |

Rationale for Choosing Primary and Foreign Keys for the Table 3:

In generating revenue, Payment is an entity of its own. For that reason a table is needed.

The Payment\_ID is the unique identifier for payment information in the “Payment” table. For that reason, Primary Key for this table. Customer\_ID and Order\_ID are both Foreign Keys in this table.

**Table 4**

**[Employee]**

|  |  |  |  |
| --- | --- | --- | --- |
| Employee\_ID [PK] | Employee\_Name[FK] | Department | Employee\_Phone |

## **Section 3:** Extraction and Transformation

Now that you have the requirements from your stakeholders, you want to understand the current state of what data is collected. That is how you recognize which additional data you need to achieve the future state. You ask the engineering team what data they are currently collecting in the pipelines and they provide you with a section\_3\_event\_logs template (which you can download from the classroom) generated by rider’s activities on the Flyber App. Also provided in the Project Resources.

**Extraction and Transformation-1**

ETL is performed on the provided Event Logs Template and results will be transferred to the proposal template. The project's ETL should be created inside of your copy of the Event Logs template in the tab titled, ETL. Clicking on the link above will create a copy of the Event Logs for you

After being provided with a CSV log file, use extraction techniques to be able to get the data into a usable form. Because this needs to be a repeatable process we need to document it in order to assess its feasibility. Below,

1. Write the steps you took to extract the data and provide reasoning for why you used this method *Note: Don't forget to include any file type changes*:
2. Perform cleaning and transformation of the data in the ETL tab and document.
3. Document and provide rationale for all of your steps below as well.

Steps for Extraction:

1. Inspect the data - Verify and understand the data entities
   1. The data should be inspected to get a clear understanding of the condition or to be familiar with the structure and issues that may exist. Does it have a consistent structure ( a CSV table) or is it unstructured (collection of email messages)?
   2. How is each data point identified. Is there an explicit, unique ID for each data point, or will one need to be created?
   3. Identifying bad data, superfluous columns (e.g., columns with only one value or missing an overwhelming amount of data), duplicate rows and things of that nature.
   4. Count how many data points or rows exist. If the data is structured, count the number of columns and missing values in each column.

Know the data type for each column, we want to know what the data type is and what the data type should be.

1. Plan
   1. After inspecting the data and having an understanding of the overall condition, there needs to be a plan of what to fix and how to fix them*.* This requires articulating the problems clearly and creating a plan to modify and/or fix the problem.
   2. With clearly stated steps to fix the problem, we can make an informed decision about whether implementing the plan is worth the effort. Sometimes there are multiple viable resolutions to choose from. To decide, we weigh trade-offs and ultimately choose the best option.
2. Execute or Implement - Clean the data
   1. Once we have a detailed list of steps to modify our dataset, it’s time to move forward with the implementation. This may involve developing the code to fix the problem being focused on.
   2. During the implementation, we might discover that the problem is more complicated than initially expected. It is anticipated that there could be unintended consequences that may arise in the process. A decision will have to be made whether to proceed or abandon the modification.
3. Inspect - Verify the data again
   1. After implementing the changes, we need to return and inspect the data in a new iteration. This step is necessary when modifying data structure, which can lead to missing data points, thereby creating more bad data.

**Transformation-2**

Analyze the data from part 1 to answer the following questions:

1. How many events are being recorded per day?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Date | 10/5/2019 | 10/6/2019 | 10/7/2019 | 10/8/2019 | 10/9/2019 | 10/10/2019 | 10/11/2019 |
| Event Count | 9891 | 18056 | 18202 | 17963 | 17600 | 17694 | 17595 |

1. How many events of each event type per day?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Date | 10/5/2019 | 10/6/2019 | 10/7/2019 | 10/8/2019 | 10/9/2019 | 10/10/2019 | 10/11/2019 |
| Choose Car | 1498 | 2843 | 2953 | 2769 | 2725 | 2801 | 2804 |
| Search | 1484 | 2891 | 2824 | 2899 | 2749 | 2904 | 2821 |
| Open | 6594 | 11733 | 11767 | 11662 | 11531 | 11325 | 11371 |
| Begin Ride | 38 | 49 | 62 | 86 | 57 | 57 | 78 |
| Request Car | 277 | 540 | 596 | 547 | 538 | 607 | 521 |

1. How many events per device type per day?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Date | 10/5/2019 | 10/6/2019 | 10/7/2019 | 10/8/2019 | 10/9/2019 | 10/10/2019 | 10/11/2019 |
| ios | 2384 | 4337 | 4217 | 4373 | 4380 | 4482 | 4500 |
| android | 1463 | 2870 | 2854 | 2729 | 2744 | 2562 | 2672 |
| Desktop Web | 895 | 2007 | 1600 | 1958 | 1712 | 1866 | 1777 |
| Mobile Web | 5149 | 8842 | 9531 | 8903 | 8764 | 8784 | 8646 |

1. How many events per page type per day?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Date | 10/5/2019 | 10/6/2019 | 10/7/2019 | 10/8/2019 | 10/9/2019 | 10/10/2019 | 10/11/2019 |
| Search Page | 3995 | 7219 | 7307 | 7221 | 6979 | 7201 | 7137 |
| Book Page | 1977 | 3548 | 3576 | 3572 | 3586 | 3424 | 3506 |
| Driver Page | 965 | 1823 | 1871 | 1794 | 1755 | 1689 | 1768 |
| Splash Page | 2954 | 5466 | 5448 | 5376 | 5280 | 5380 | 5184 |

1. How many events for each location per day?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Date | 10/5/2019 | 10/6/2019 | 10/7/2019 | 10/8/2019 | 10/9/2019 | 10/10/2019 | 10/11/2019 |
| Manhattan | 6869 | 12591 | 12807 | 12180 | 1220 | 12371 | 12201 |
| Brooklyn | 2009 | 3737 | 3590 | 4025 | 3440 | 3400 | 3556 |
| Bronx | 250 | 533 | 507 | 469 | 510 | 394 | 558 |
| Queens | 595 | 842 | 905 | 893 | 1026 | 1069 | 936 |
| Staten Island | 158 | 363 | 393 | 396 | 354 | 460 | 344 |

**ETL Automation and Scalability:**

Provide an analysis about this ETL process. Address and provide rationale for manually extracting, loading and transforming the data from the raw logs. Also address potential preliminary recommendations on improving this process.

* Background

*Raw data was provided in an Excel CSV format in the file “section-3-event-logs-template”. The file consists of 9 Columns of 124981 records of events of different types, from various sources, as well as, different timestamps. The data is somewhat organized, but not properly structured for easy analysis.*

*The ETL process allows extraction of the data from the raw file, which is now rearranged and transformed into a scalable and segmented structure that is easily understood for analysis. We can view the events based on Event Type, Daily Volume, Device Type, Page Type and Location. The transformed data presents a better opportunity to utilize the data to its full potential in any application of interest. Note that ETL process was done manually in this case, partly due to very small size data.*

* *Advantages of Automated ETL Process*

Working with the automated ETL process helps in time saving. Also, you will be able to initiate many processes and engage automatic extraction which is normally done otherwise.

Working with Automated ETL helps to adapt rapidly changing data. Also, we can easily handle a huge amount of data as compared to the manual extraction that requires us to work with it bit by bit hence tedious.

Note that manual ETL will not be suitable for a large amount of data, because it will require working on small bits of the data sequentially, which takes much longer.

* *Tool for Analysis*

*Tableau provides a major advantage as a tool for use for analyzing, planning and organizing data in the process. Depending on the size and condition of the raw data, it could be more complicated or much more difficult.*

* Recommendation

*Automating the ETL process will be ideal for such a process that is required to be performed periodically as the data grows. T*he manual process becomes more tedious and time consuming as previously highlighted.

Was the process of manually Extracting, Loading and Transforming data from raw logs efficient?

No, it is not efficient. Automated ETL process is needed.

Is it scalable?

No, not scalable

Is there a need for an automated ETL pipeline?

Yes.

Why?

ETL is a required process periodically for so many reasons as data changes and grows in size.

**Section 4:** Choosing Relevant Dataset

The previous exercise gave you a sneak peek into the Extraction and Loading aspects of ETLs in data pipelines. For making business decisions, a data consumer would like to have all the data they want. However, for any ecosystem, it is impossible to collect or provide everything that the customers need. In this exercise, you will get a taste of real world scenarios wherein:

* All the resources are not always available to get what you need.
* You have to get creative and get the most insights with a minimal data set.

Oftentimes your stakeholders/customers will “ask for the moon”, but you’ll have to push them to work with the small amount of information you have and get creative.

***Note: As you learned in the course, being a Data Project Manager involves an extraordinary amount of collaboration. Complete the next sections based on the following scenario.***

After the analysis in section 3, we made sense of the numbers, and realized the total number of events seems to be too small (this was a week's worth of data, but you need at least a month). Further investigation reveals that this was a subset of logs, but the actual data that is being collected is much bigger. Working through this small data set was tedious, and repeating this exercise on a much bigger data set manually won’t be feasible. Considering the time constraints of this project, engineering is willing to help with some automation. They also have limited bandwidth and are busy scaling systems up.

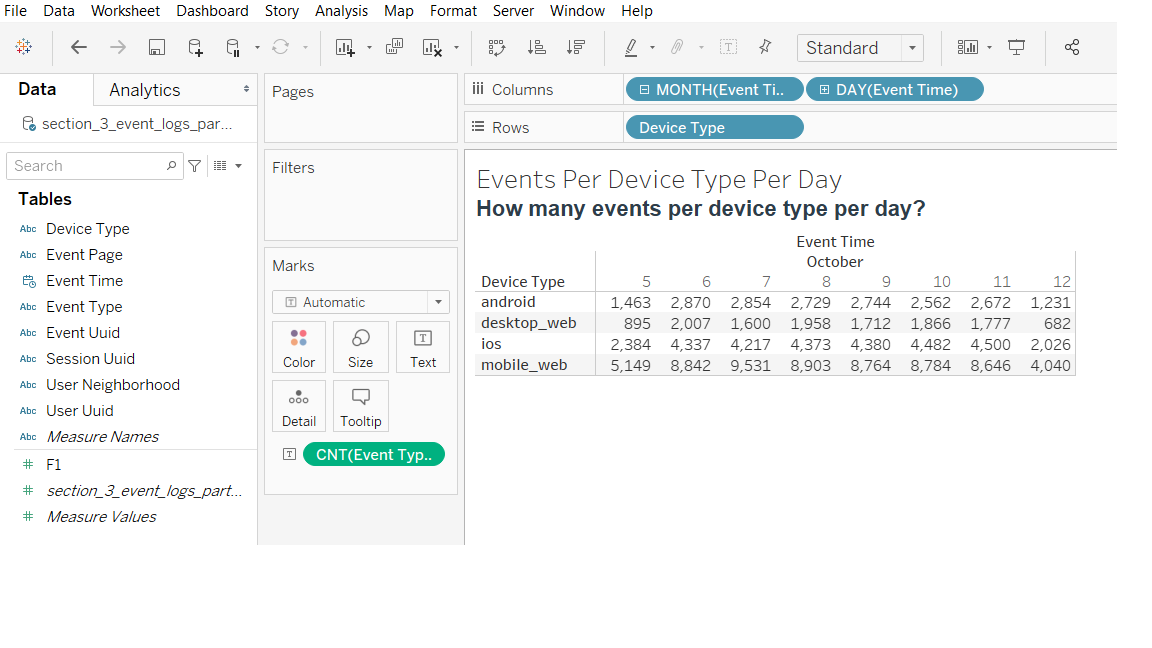
Engineering is willing to provide some data, but they have asked for the criterion that is most important. To First provide your business question and provide a rationale for why this is the most important.

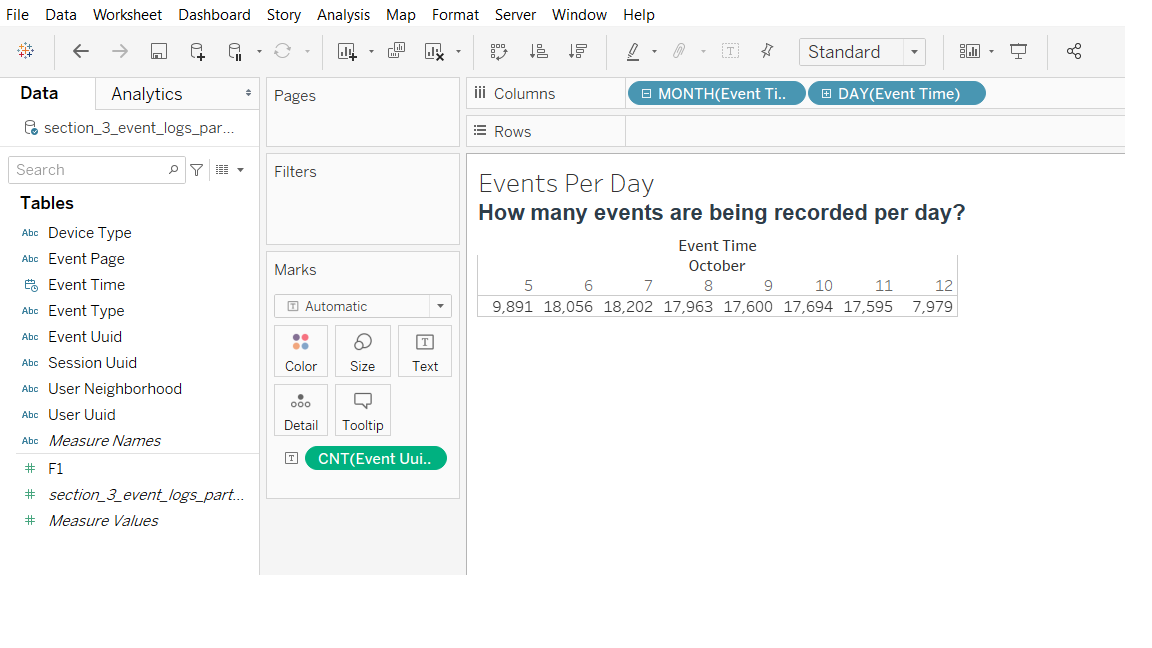
Choose one of the following prompts that you think can get you the most relevant information to proceed further.

1. How many events are being recorded per day?
2. How many events of each event type per day?
3. How many events per device type per day?
4. How many events per page type per day?
5. How many events for each location per day?

For your chosen question also answer the following using the data from section 3 to support your answer:

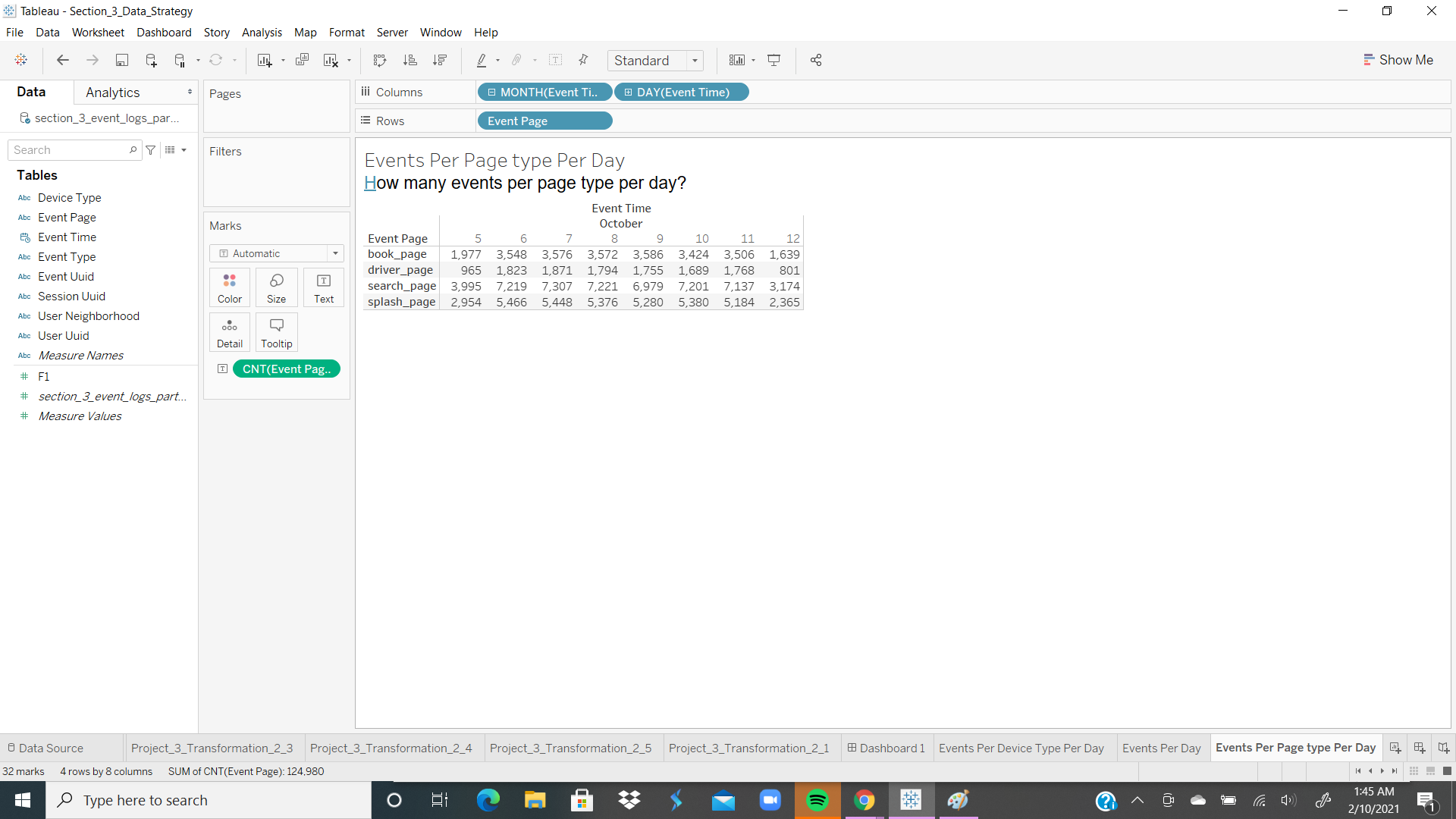
1. How much is the customer data increasing?
2. How much is the transactional data increasing?
3. How much is the event log data increasing?





Which of the following data is ***most*** important to answer this question? Why?

* Event Log Data
* Transactional Data
* Customer Data

**

How many events per page type per day?

Event Log Data is the most important to answer the question.

The reason being the fact that the Event Log consists of a representation of most aspects of the data in some form. There may be a need for extraction.

## **Section 5: [Optional]** Loading and Visualization On Your Own This sectional is an optional part of the project that you can do to make it standout. We have provided visualizations in the appendix if you decide not to do this section. You can also use our visualizations to compare what you created

## After sharing your criterion with engineering, they give you a new set of data: Section 5 Event Type Log also available in the classroom resources. Also provided in the project resources section.

Engineering provided you with the data you want, but you still have yet to achieve your ultimate goal as a Data Product Manager. Now, utilize the data to make business decisions. Your executives do not want you to give them a bunch of data tables; instead, they prefer visualizations to help convey the key insights succinctly. Visualizing this data will help you understand the underlying trends and help you determine the story that needs to be told in your proposal to executives.

In this section, you can load and visualize the data into whatever platform you would like. A Python Notebook, Tableau or any other visualization tool you are familiar with. Create two visualizations that might help you to better understand your data trends and place either a screenshot or exported image of your visualizations and the details of each below. Please provide the steps you took to visualize your data and what the visualization tells you about your data.

Visualization 1:

*[Insert Visualization Here.]*

**Data Story:** This graph tells us:

*[Insert Response Here.]*

This graph was created using the following steps:

1. *[Insert Step Here.]*
2. *[Insert Step Here.]*
3. *[Insert Step Here.]*

Visualization 2:

*[Insert Visualization Here.]*

**Data Story:** This graph tells us:

*[Insert Response Here.]*

This graph was created using the following steps:

1. *[Insert Step Here.]*
2. *[Insert Step Here.]*
3. *[Insert Step Here.]*

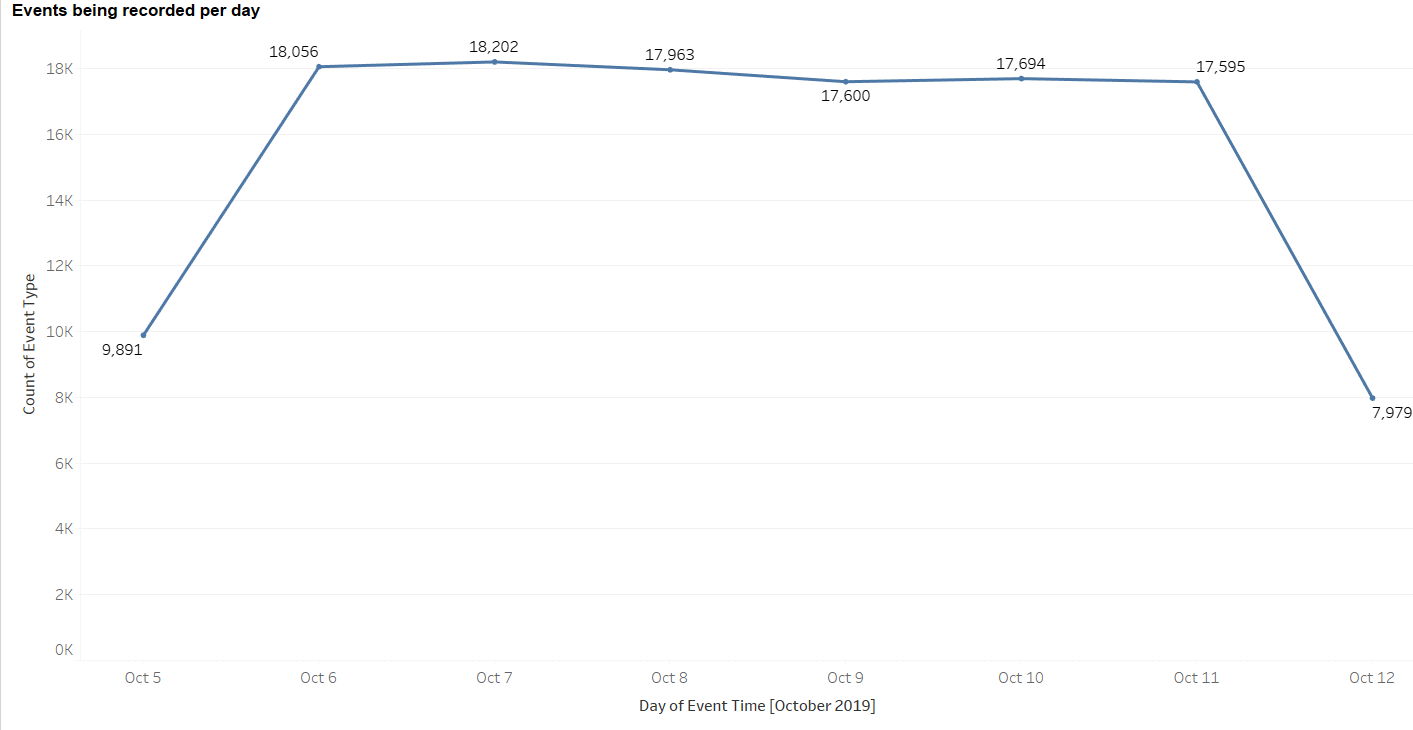
## **Section 6:** Business Insights

The Data is loaded and ready for analysis. We want to use this data as evidence to support our recommendations. It is important that we understand this data and the underlying trends and nuances that these visualizations show us. As you already know, any proposal backed up by data is always better received and considered more robust.

What is the story the data is telling you about Flyber's data growth? If you created Visualizations, you can use them as well, but they are not required). Include any data and calculations that were made to help tell that story and quantify the data growth.

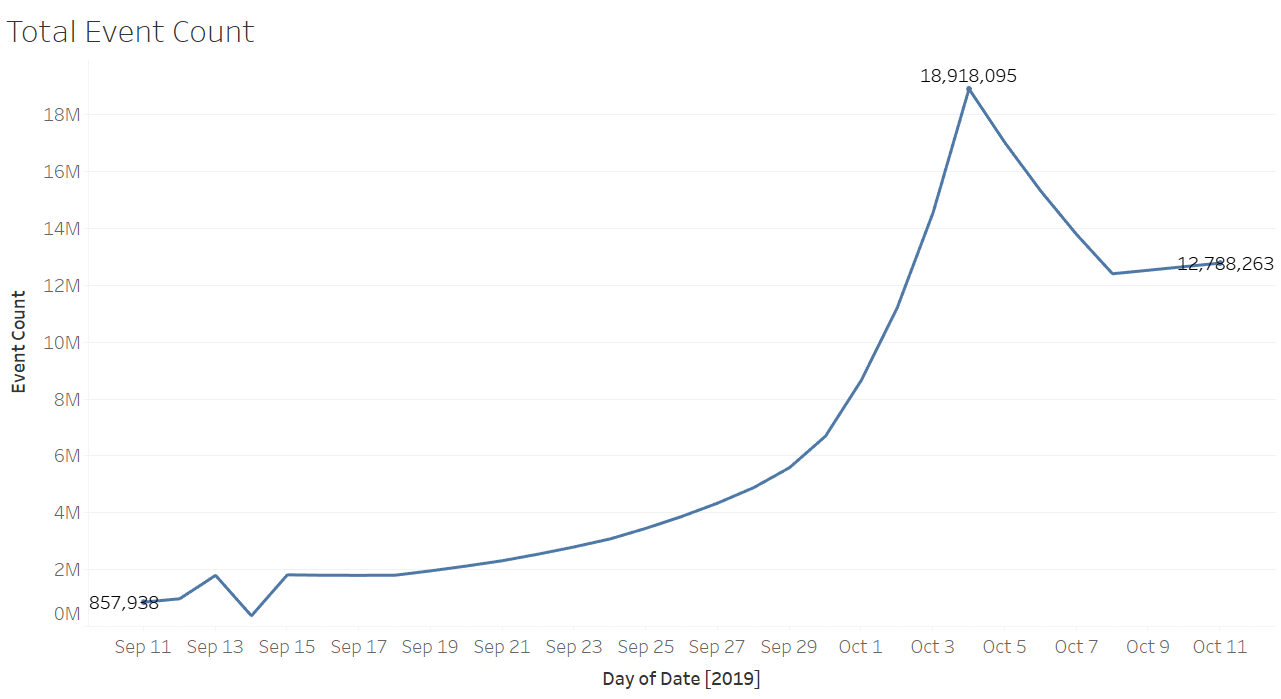
**Data Growth for Last Month**

Visualization:



*[Insert Visualization Here] If you didn't create your own, please use the images in the appendix to guide you through this section.*

Data and calculations used for quantifying of Flyber's Data Growth:



Growth (g) = percentage change in data between the two months

Growth(g) = ((Final value in Oct - Initial value in Sep) / Initial value in Sep) \* 100

Initial value in Sep = 790,329

Final value in Oct = 12,788,264

Growth(g) = ((12,788,264 - 790,329) / 790,329) \* 100 = 1518%

Note: I decided to use the graph in the appendix because the data provided in the raw data was for the month of October 2019 only. Data for the previous month was not included. However, the appendix offered that.

*Looking at the Event Log data, there is a sharp increase of 8,165 from Oct. 5 to Oct. 6, 2019.*

*Volume remains steady through Oct. 7. From there, minor decrease was consistent through Oct. 11, 2019, which was followed by a sharp decline of 9,616 down to a volume of 7,979.*

*Event volume was highest on Sunday and Monday. Slight decrease on Tuesday, Weds., Thursday and Friday. Saturdays have the lowest volume of events.*

*Insert Response Here.]*

What is the fastest growing data and why?

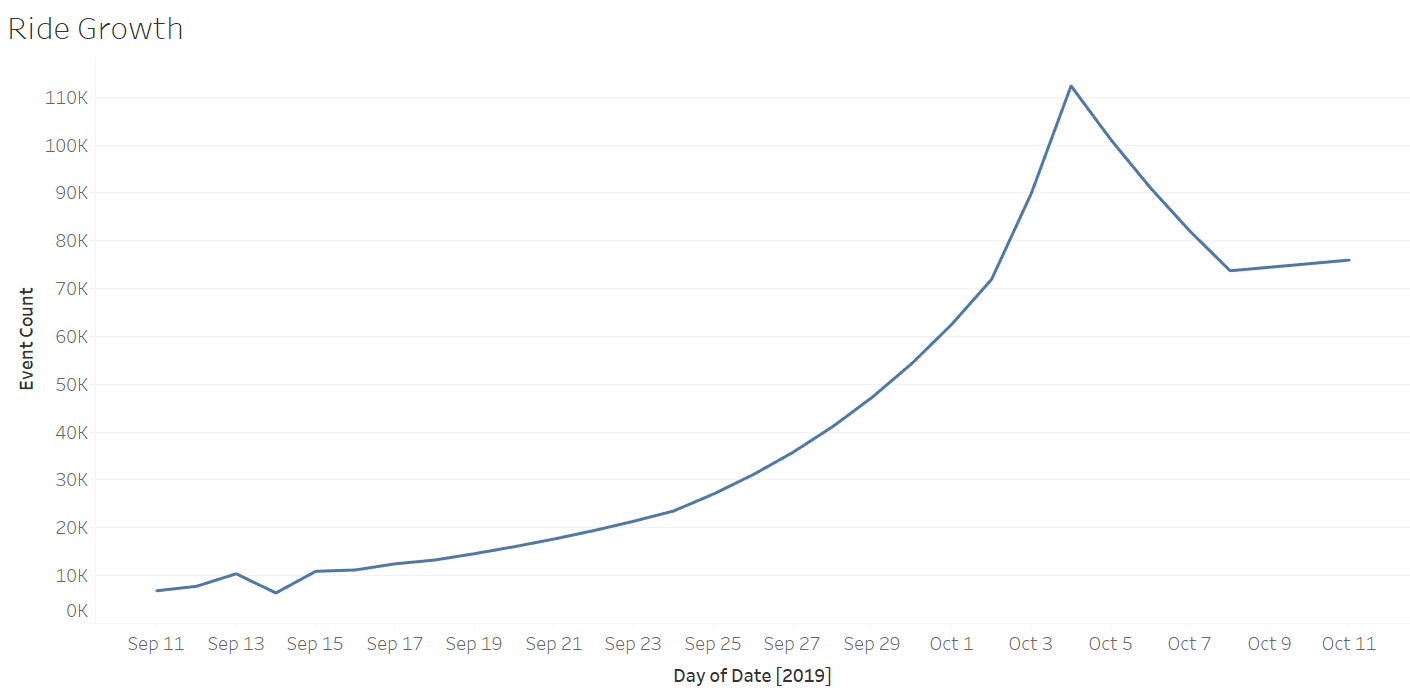
Mobile Web is the fastest growing data

The data showed consistent growth throughout the period. It is a confirmation of the fact that users are more engaged online with mobility.

*Insert Response Here.]*

**All Event Type Data**

Visualization:



*[Insert Visualization Here] If you didn't create your own, please use the images in the appendix to guide you through this section.*

What is the Data Story our data tells for each of the following:

* Graph Pattern
* Good or Bad
* October Marketing Campaign
* Marketing Campaign Impact
* Importance of Relationship Between Marketing Campaigns and Data Generation
* Graph Pattern

Ride Growth showed consistent increase from September 11 and peaked on Oct. 3/4. Then, followed by a dip on Oct 8.

* Good or Bad

The indication is Good based on the result and contribution to Ride Growth.

* October Marketing Campaign

The October Marketing Campaign yielded higher increases in Ride Growth.

* Marketing Campaign Impact

The impact of marketing campaigns is significantly important in generating data. Increased advertising is needed during the month of October to maintain upward trend Ride Growth.

* Importance of Relationship Between Marketing Campaigns and Data Generation

Users need to be effectively engaged during advertising, That increases user retention rate, which in turn increases data generation.

## **Section 7:** Data Infrastructure Strategy

Thus far we have:

* identified data stakeholders and their data needs.
* Identified what data is currently being collected and what data needs to be collected.
* Identified data insights and growth trends.

Now, it's time to tie all the loose threads together and bring this process to its logical conclusion by suggesting which Data Warehouse (DWH) Flyber should invest in and why. Using data warehouse options below, suggest whether Flyber should choose an on-premise or Cloud data warehouse system and which specific data warehouse would best serve Flyber’s data needs.

**Data Warehouse Options**:

Cloud:

* Amazon Redshift
* Google BigQuery
* Snowflake
* Microsoft Azure

On-Premise:

* Oracle Exadata
* Teradata, Vertica
* Apache
* Hadoop

You will address the following factors with a rationale as to why the DWH chosen is the best for Flyber:

* Cost
* Scalability
* In-house Expertise
* Latency/Connectivity
* Reliability

**Cloud vs On-Premise**

Provide an evidence based solution as to why Flyber would be best served by a Cloud or on-premise DWH. In this response, you don’t need to specify *which* specific Cloud or on-premise DWH product you will choose, just if it will be Cloud or on-premise. Remember to address the factors above.

* Flyber will be best served with a Cloud based system.

Cost - There is no need to purchase and maintain any hardware for the system. Significant

cost saving advantage. It’s quicker and cheaper to set up Cloud Data Warehouse system

Scalability - easy to scale up or down as needed. Demand tends to change seasonally, weekly, and hourly. The Consequences of not having enough compute or storage resources are dire. First come performance issues, then users start getting error messages and getting locked out of the application. The Cloud architecture prevents all of these.

In-house Expertise - Flyber company should focus on its core area of expertise and outsource Cloud services to the experts on that. Developing in-house expertise tends to take more time and resources.

Latency/Connectivity - Data is needed for real-time use-cases. Speed is an advantage with the Cloud. Cloud-based data warehouse architectures can typically perform complex analytical queries much faster because they use massively parallel processing (MPP)

Reliability - Having a Cloud based system is more reliable. Accessing an app or service in the cloud by so many users and so many resources can be demanding. With the Cloud, you can reasonably expect that the app or service is up and running. You can access what you need from any device at any time from any location. There will be no interruptions or downtime.

*Insert Response Here.]*

**Suggested DWH**

Provide an evidence based solution as to which DWH product is best for Flyber. Remember to address the factors above.

* Amazon Redshift will be most suitable for Flyber Flying Taxi business

We have to ensure that the choice we make meets the requirements of the Flyber now and in the future.

Amazon Redshift has what is known as RA3 nodes, which allows you to scale, compute and only cache the data you need locally. You cannot separate workloads. While you can only run Redshift as an isolated workload on AWS, it has the most options on AWS, including the ability to deploy it in your own Virtual Private Data Center.

Cost - Redshift is considerably cheaper than alternatives or an on-premise solution. Redshift has two pricing models that give you the flexibility to categorize the expense as an operational expense or capital expense.

Scalability – Scalability is crucial for any data warehousing solution, and Redshift performs well in this arena too. It is horizontally scalable, meaning whenever you need to increase the storage or need it to run faster, you can add more nodes using AWS console or Cluster API, and it will upscale immediately.

Latency/Connectivity — When it comes to loading data and querying it for analytics and reporting, Redshift is extremely fast. MMP allows you to load data at fast speeds. The Redshift Query Engine has the same interface as PostgreSQL, which means developers who are already familiar with SQL won’t have a steep learning curve to get going. Since Redshift uses SQL, easily connecting to other business intelligence tools.

Security — Redshift comes packed with security features, including various ways to handle access control, Virtual Private Cloud (VPC) for network isolation, data encryption etc... You can launch Redshift clusters inside your VPC so you can define security groups and restrict inbound or outbound

*Supporting research resources*

[*Snowflake vs Amazon Redshift*](https://hi.firebolt.io/lp/snowflake-vs-redshift-2v?utm_term=amazon%20redshift%20vs%20snowflake&utm_campaign=competitors-snow&utm_source=google&utm_medium=cpc&utm_content=491237745516&hsa_acc=1333801409&hsa_cam=11142868968&hsa_grp=111993189569&hsa_ad=491237745516&hsa_src=g&hsa_tgt=kwd-941339898864&hsa_kw=amazon%20redshift%20vs%20snowflake&hsa_mt=p&hsa_net=adwords&hsa_ver=3&gclid=Cj0KCQiAgomBBhDXARIsAFNyUqPERm8iJeGK_Opa7rMHHZ5MSGaoCbFLYyo-EPNqugaBJRop1Ns6HJQaAujeEALw_wcB)

[*Pros and Cons of Amazon Redshift*](https://www.verypossible.com/insights/data-warehouse-showdown-pros-and-cons-of-amazon-redshift)

*Insert Response Here.]*

## Image Appendix

Image 1: Log Growth

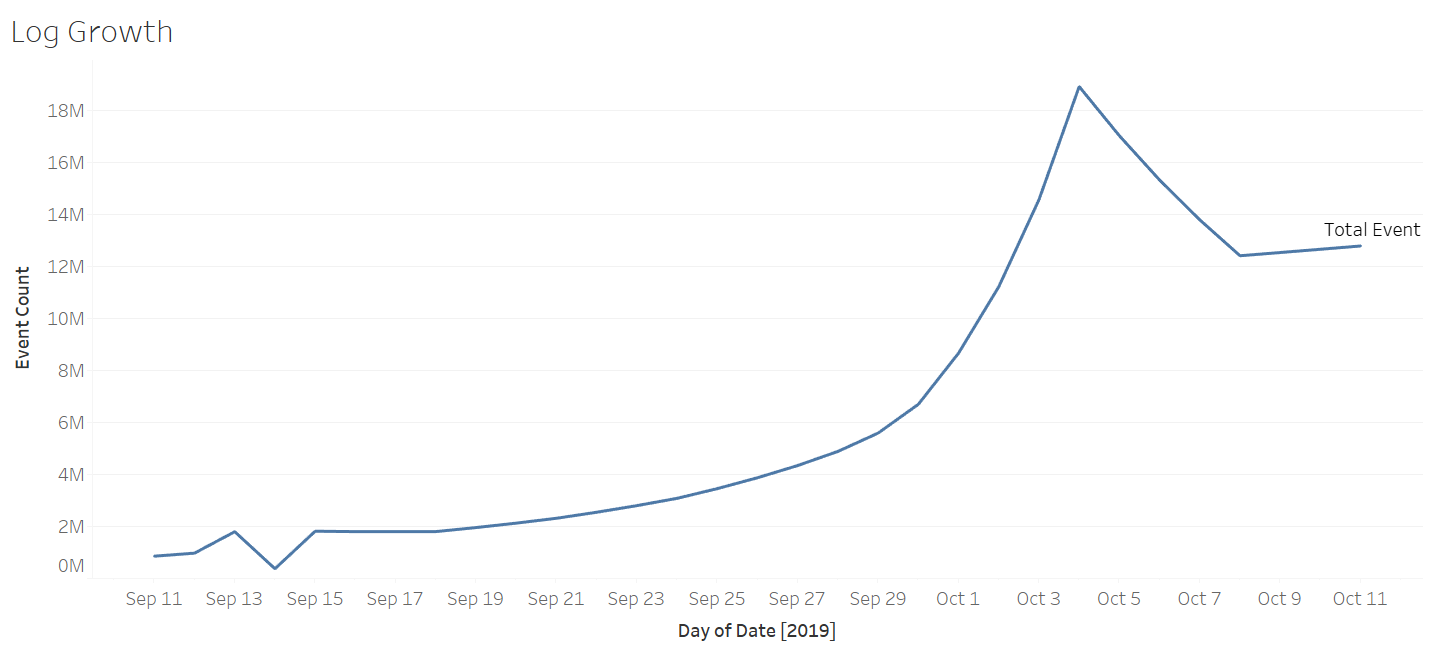


Image 2: Ride Growth

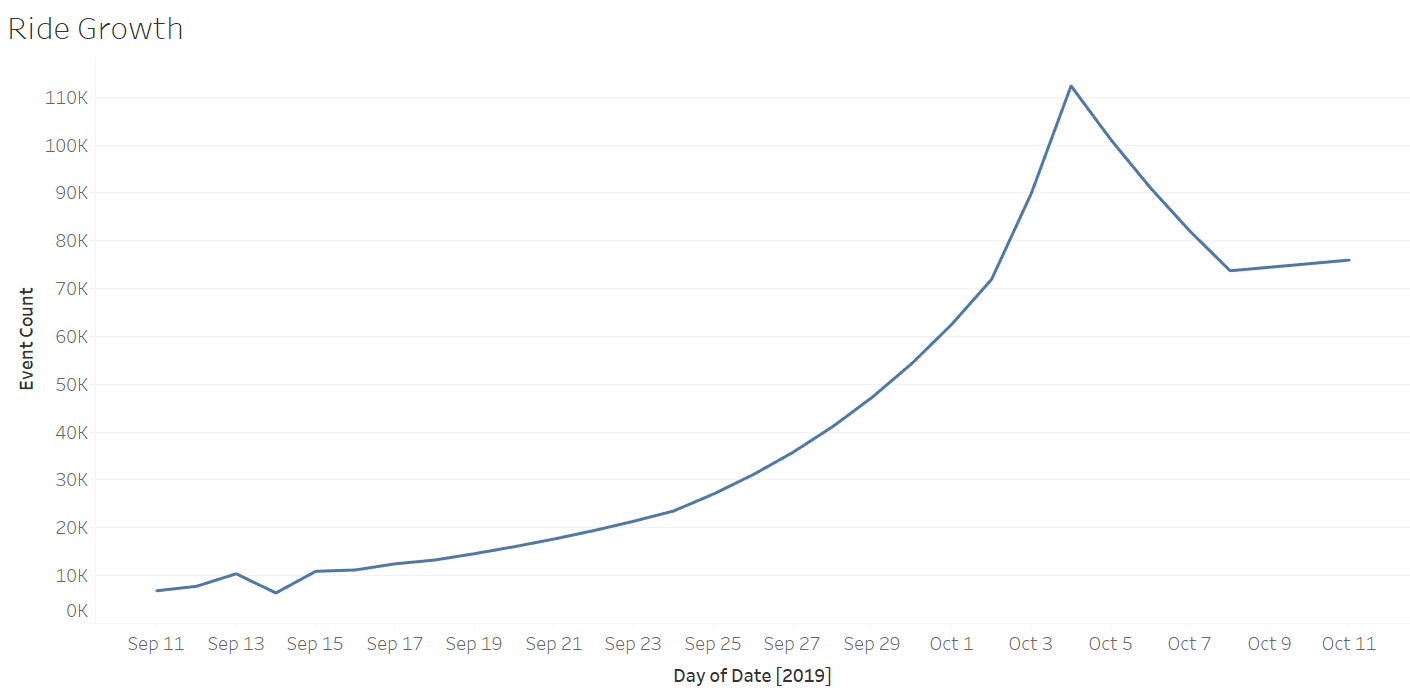


Image 3: Total Event Count

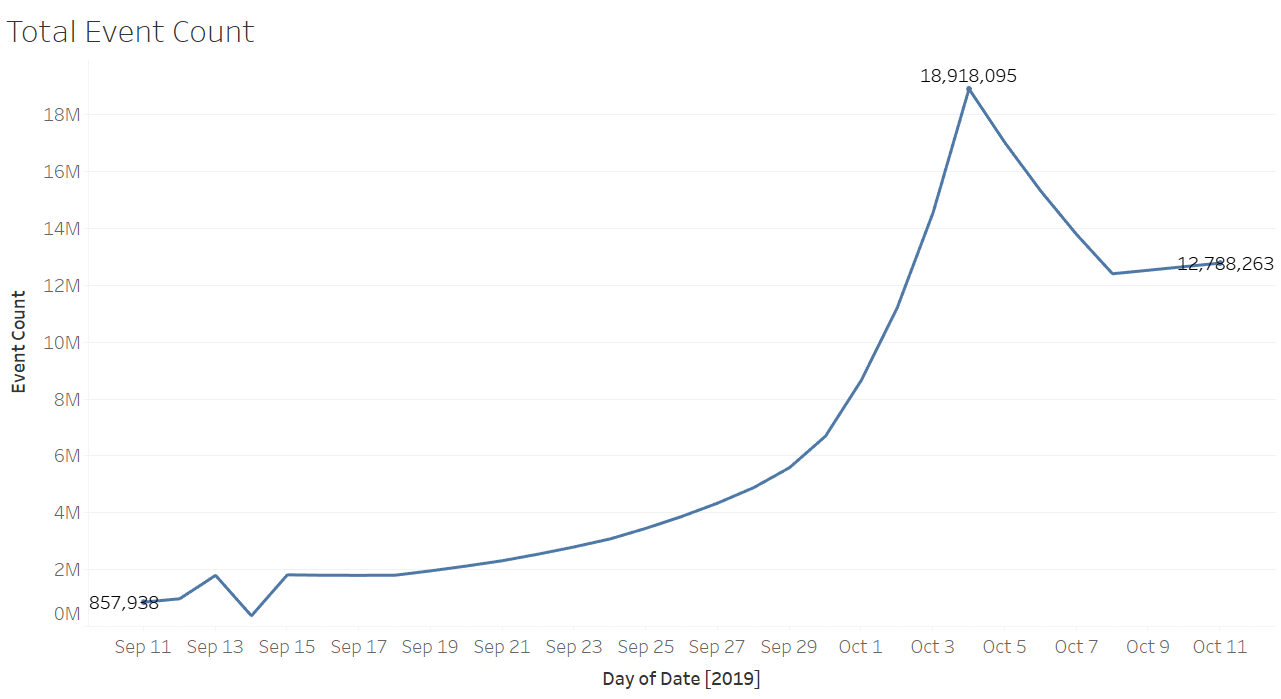


Image 4: All Events Log Scale

