# Report

**Airlift Technologies**

**Supply Data Scientist-- Case Study**

**Name:** Syed Arham Ali Rizvi

**Date:** 12/08/2021

**Specific Instructions for the Case Study:**

**Context --** Airlift's vision is to grow into a household name and remain top of mind for customers with an effective marketing strategy. As the business scales and grows into new geographies, we will need to have targeted customer communications, precise estimates of the impact of our growth levers and data driven understanding of how we can improve our product.

**Background**: Currently Airlift has 3 different products ranging in 3 different industries (This information can be found on their website: https://www.airlifttech.com/):

1. Airlift Express (Grocery Delivery Service)
2. Airlift Transit (Commuting Service)
3. Airlift Bookshelf (Books/Education)

**The airlift dataset was used for analysis.**

**All the analysis was done using Jupyter Notebook, Python, Pandas and Matplotlib.**

**The 4 notebooks can be found in the same folder as this report.*

**Question 1:**

We are struggling with new user acquisition -- what are different data points and strategies you would use to push acquisition across markets?

**Answer:**

There are several ways to **acquire new customers**, here are a few:

**1. Content marketing**: publish content via blog posts, ebooks, guides, videos, etc. Emailing & social media can be used to promote it. It also generates about 3x as many leads as other marketing methods, and costs 62% less

**2. Highly targeted advertising**: study your audience & target them. Market segmentation may help.

**3. Developing business partnerships**: Partner up with well established businesses that may help your business. For example, Microsoft.

**4. Create a lead generating site**: It is important to have an attractive, professional website that will grab the users attention thus generating leads.

**5. Focus on benefits over features**: List the features briefly, but focus primarily on the benefits.

**6. Be present on social media**: Statistics show that 97% of adults (ages 16-64) log into at least one social network per month. Platforms like Facebook and Instagram also offer advertisements and market-places at a fairly cheap rate.

**7. Make your brand known on forums**: Answer questions on forums such as Quora.

**8. Offer deals and promotions**: People love discounts because they save money and feel like they're getting access to something exclusive and limited.

**9. Run giveaways**: If people aren't using your product, consider giving gifts or merchandise. This not only establishes a good connection with the customers but also serves as marketing.

**10. Showcase testimonials**: One of the best ways to acquire new customers is to highlight existing ones. Good reviews show reliability.

**11. Keep track of the competition**: Actively monitoring competitor's marketing tactics, backlinks and traffic, web design, social mentions, products or services, etc. is the best way to stay ahead of the game.

**12. Host an event**: Hosting an event is a great way to do this as it gives potential customers the opportunity to meet your brand on a more personal level.

**13. Ask for referrals**: Use existing customers to get new customers by offering incentives.

**14. Make sure your SEO is up to date**: SEO complements content marketing efforts by optimizing content and making it easier for your target audience to find it.

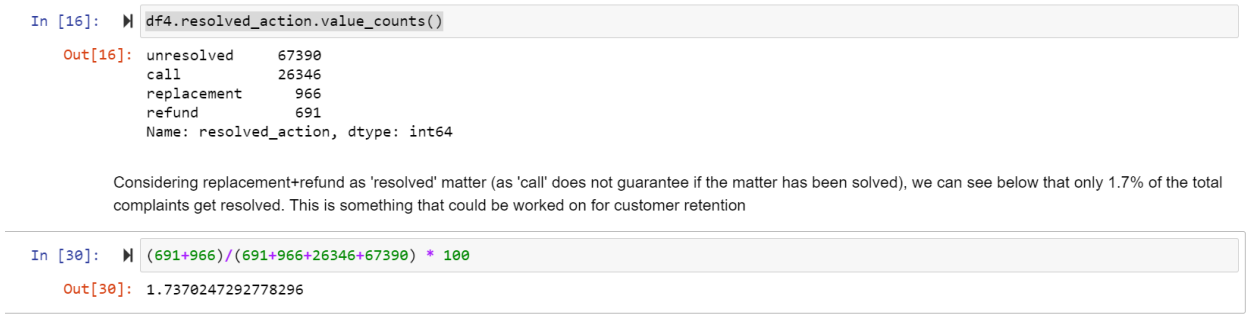
**15. A/B test everything**: A/B testing is crucial in determining which strategies and campaign components produce the highest numbers of conversions.

Now Considering possibilities based on the Airlift-technologies 3 services:

- **Airlift Transit** can expand their operations in more major cities and include inter-district/long travels.
- **Airlift Transit** can launch tour buses. Since tourism is on a rise in Pakistan, Airlift express could benefit greatly by providing a cheaper and reliable service.
- **Airlift Express** can provide unique items that aren't readily available in the markets, For example, 'Springs grocers' have supplies to items that aren't available in most of the other stores.
- **Airlift Express** can provide scheduled grocery service like 'Tesco'.
- **Airlift Express** can expand from just delivering groceries to delivering any package for example, take the case of 'Bykea'.

Based on the **Datasets** provided:

1. There were a substantial number of complaints which remained 'unsolved'. That can be detrimental to a company's reputation and competitors like 'PandaMart, Foodpanda' can take advantage of the situation and try to gain your customers. Customer retention should be a priority. (The analysis can be seen under **'Airlift_Technologies_dataset4.ipynb'**)

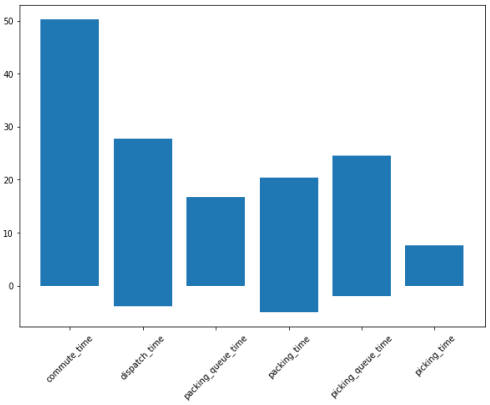


```
In [16]: df4.resolved_action.value_counts()

Out[16]: unresolved     67390
         call           26346
         replacement      966
         refund           691
         Name: resolved_action, dtype: int64
```

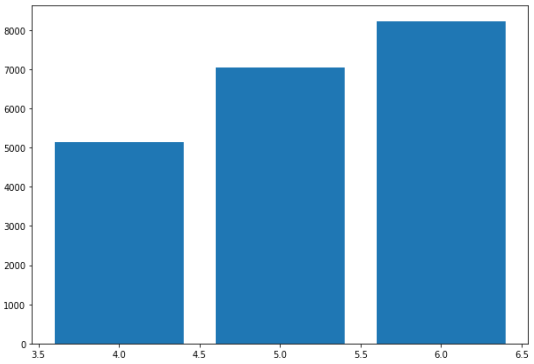
Considering replacement+refund as 'resolved' matter (as 'call' does not guarantee if the matter has been solved), we can see below that only 1.7% of the total complaints get resolved. This is something that could be worked on for customer retention

```
In [30]: (691+966)/(691+966+26346+67390) * 100

Out[30]: 1.7370247292778296
```

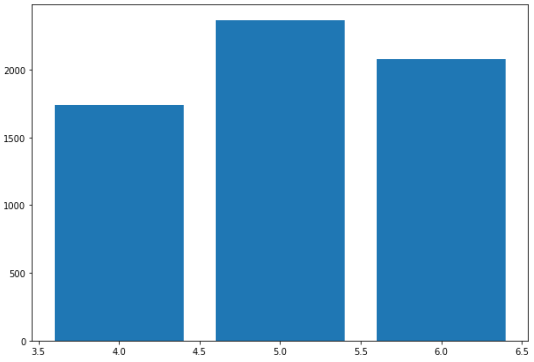
2. The dataset also showed that the most amount of time was taken by 'Commute time' followed by 'dispatch time'. Improvements can be made in these departments to provide the customer with the package earliest possible. (The analysis can be seen under **'Airlift_Technologies_dataset3.ipynb'**)



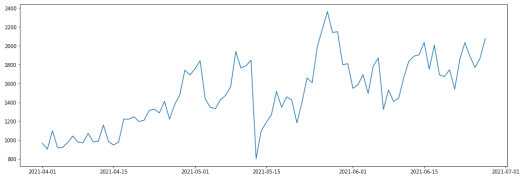

3. **From data-set1**: Although the overall orders increased from April 2021 - June 2021 (from approx. 5000 - 8000 orders), grouping the orders by ‘New’ and ‘Repeat’ showed slight variations. The ‘New’ order types saw a dip in the number of orders from May 2021 - June 2021. While the ‘Repeat’ order types saw a steady increase from April 2021 - June 2021. (See images below) (The analysis can be seen under **‘Airlift_Technologies_dataset1.ipynb’** ):



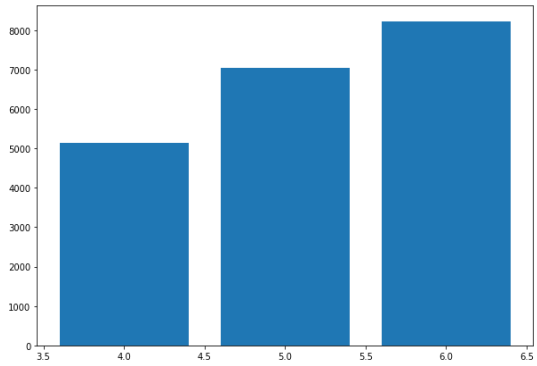

Overall



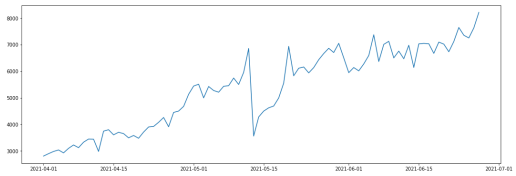

Order_type = ‘New’



Order_type = New - time series




Order_type = Repeat'



Order_type = Repeat - time series

**Question 2:**

We want to increase our average order value (cart size) -- what data points can we study to determine how we can do this? What ideas do you have in terms of what we can incorporate into our product?

**Answer:**

Following are a few ways to increase the average order value (cart size):

1. **Incentivize minimum spend**: Offer deals such as '15% off coupon for baskets containing three or more items'
2. **Recommend products**: include related products. Include comparisons on product pages of more expensive items, including links, photographs, and customer reviews as to why the more expensive option is better. Enable customers to "quick add" complementary products to their basket at the checkout stage. Add a banner of products that "other customers also bought" to product and category pages.
3. **Increase perceived value**: use consumer psychology to increase the perceived value of goods. Examples include: Bundled products, Price anchoring, Free shipping
4. **Build a relationship**: Customers tend to spend more on Companies they trust. Include Loyalty programs, Recommending items on email, Live chat etc.

Now Considering possibilities based on the Airlift-technologies 3 services:

- **Airlift Transit** can expand their operations in more major cities and include inter-district/long travels.
- **Airlift Transit** can launch tour buses. Since tourism is on a rise in Pakistan, Airlift express could benefit greatly by providing a cheaper and reliable service.
- **Airlift Express** can provide unique items that aren't readily available in the markets, For example, 'Springs grocers' have supplies to items that aren't available in most of the other stores.
- **Airlift Express** can provide scheduled grocery service like 'Tesco'.

- **Airlift Express** can expand from just delivering groceries to delivering any package for example, take the case of ‘Bykea’.

**Question 3:**

We want to define “Retention” and “Churned Customer” -- how would you use data to form these definitions?

In simple words, **Customer Retention** refers to “how to keep a customer after acquiring him/her”. Here are a few statistics on Customer Retention:

1. The cost of acquiring new customers vs retaining them is 5 times more
2. Increasing customer retention by 5% can increase revenue by 25-95%
3. The probability of selling to existing customers and new customers is 60-70% and 5-20%, respectively
4. 90% of satisfied consumers will purchase again

**Customer retention** strategies will vary based on business model, audience, resources, and more, but here are several key qualities all of them should have:

1. Convenience
2. Altruism
3. Personalization

A **Churned Customer** is a customer that has stopped/paused using your service/product. Customer churn can be calculated by the following formula:

Churn rate = (no. of customers lost during the time) / (number of customers you had at the beginning of that time period)

Based on the **Datasets** provided:

1. There were a substantial number of complaints which remained ‘unsolved’. That can be detrimental to a company’s reputation and competitors like ‘PandaMart, Foodpanda’ can take advantage of the situation and try to gain your customers. **Customer retention** should be a priority. (The analysis can be seen under

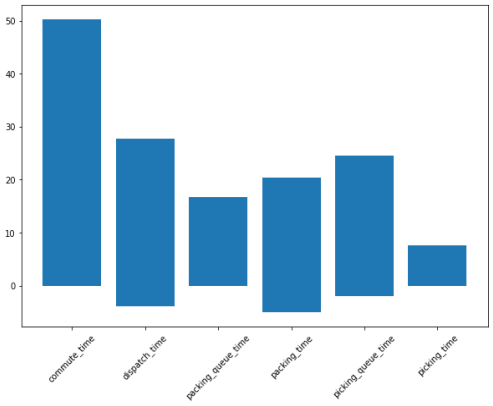
**‘Airlift_Technologies_dataset4.ipynb’**)

```
In [16]: df4.resolved_action.value_counts()

Out[16]: unresolved     67390
         call           26346
         replacement      966
         refund           691
         Name: resolved_action, dtype: int64
```

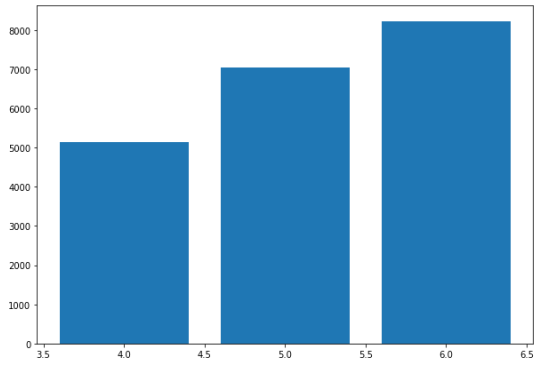
Considering replacement+refund as 'resolved' matter (as 'call' does not guarantee if the matter has been solved), we can see below that only 1.7% of the total complaints get resolved. This is something that could be worked on for customer retention

```
In [30]: (691+966)/(691+966+26346+67390) * 100

Out[30]: 1.7370247292778296
```

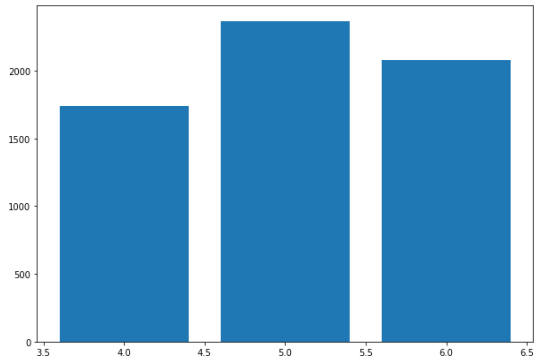
2. The dataset also showed that the most amount of time was taken by ‘Commute time’ followed by ‘dispatch time’. Improvements can be made in these departments to provide the customer with the package earliest possible. This could add on to the **‘Churned Customer Rate’** as the displeased customer might not order from your service the next time. (The analysis can be seen under **‘Airlift_Technologies_dataset3.ipynb’**)

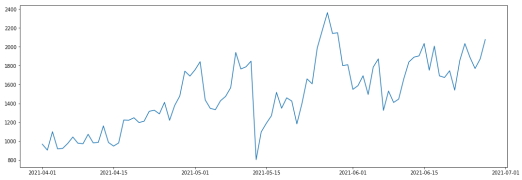



3. **From data-set1**: Although the overall orders increased from April 2021 - June 2021 (from approx. 5000 - 8000 orders), grouping the orders by ‘New’ and ‘Repeat’ showed slight variations. The ‘New’ order types saw a dip in the number of orders from May 2021 - June 2021. While the ‘Repeat’ order types saw a steady increase from April 2021 - June 2021. (See images below) (The analysis can be seen under **‘Airlift_Technologies_dataset1.ipynb’** ):

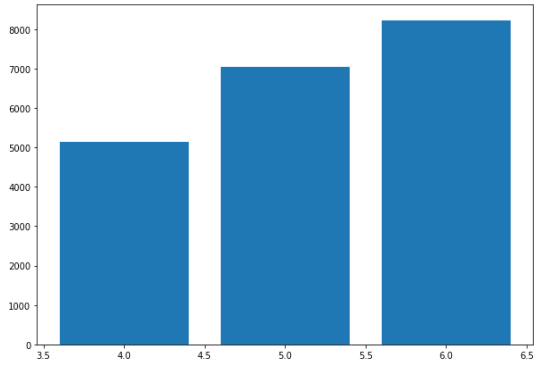


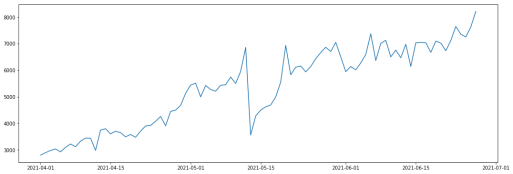

Overall (Shows good growth)



Order_type = ‘New’ (Barplot shows steady decline in new customers)  Order_type = New - time series



Order_type = Repeat’  (Barplot shows good customer retention)
Order_type = Repeat - time series

Reasons for Customer Churning can also be found via customer complaints. The analysis can be found under **‘Airlift_Technologies_dataset4.ipynb’ (complaints).**

While there are a number of complaints a customer could have, analysis showed that the major Reasons were ‘Product Quality/Damage’ followed by ‘Delivery Issues’. Most of the complaints were in May 2021. The ‘Delivery Issues’ can be further linked to Commute time problems. Please see the figures below:

**Most frequent Complaint Reason**

```
df4.complaint_reason_title.unique()
```

```
:1]: array(['Payment', 'Technical', nan, 'Product Quality/Damage', 'Delivery',
       'default'], dtype=object)
```

```
df4.complaint_reason_title.nunique()
```

```
:2]: 5
```

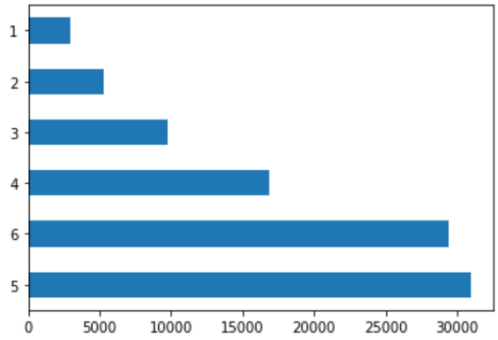
```
df4.complaint_reason_title.value_counts()
```

```
:5]: Product Quality/Damage    36334
     Delivery                  21813
     Payment                   11389
     Technical                 11026
     default                     305
     Name: complaint_reason_title, dtype: int64
```

### Months with the most Complaints

```
df4.month.value_counts().plot(kind='barh')
```

```
39]: <AxesSubplot:>
```




**Question 4:**

While we have experienced exponential growth, there came a point when we saw a significant drop in our sales volumes. You are tasked with:

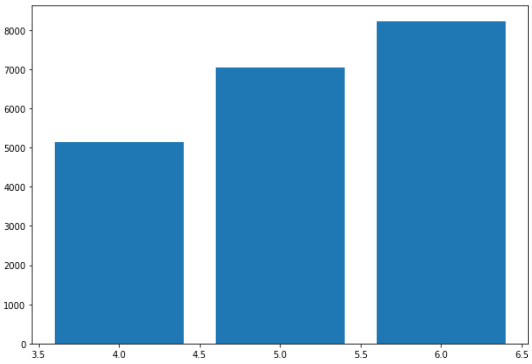
1) Figuring out when this major drop in growth occurred (**short answer: Near 15th of May 2021**)

2) Determining the major drivers behind this drop in growth (**short answer: Possibile drivers include: Product Quality/Damage, Delivery Issues, Payment Issues, Technical Issues, Warehouse Problems**). *See Answer 2 below for more detail*

**Answer1:**

Based on the **Datasets** provided:

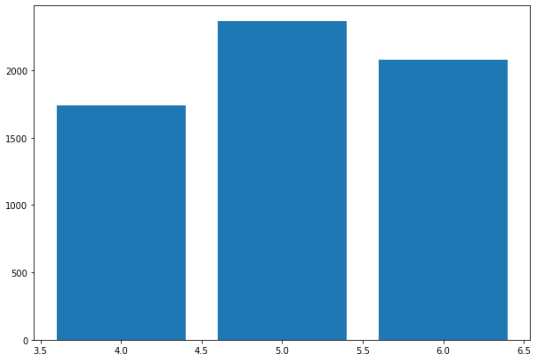
1. **From data-set1**: Although the overall orders increased from April 2021 - June 2021 (from approx. 5000 - 8000 orders), grouping the orders by 'New' and 'Repeat' showed slight variations. (The analysis can be seen under **'Airlift_Technologies_dataset1.ipynb'** ):



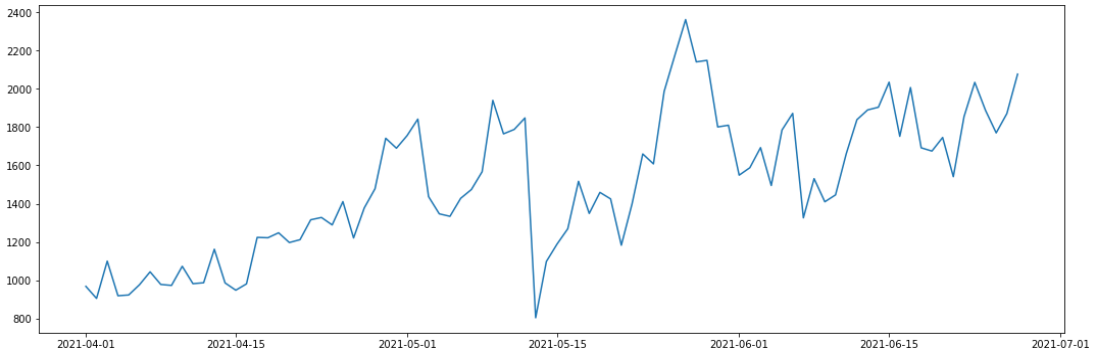

Overall (Shows positive sales)

2. The 'New' order types (barplot) saw a dip in the number of orders from May 2021 -

June 2021. The ‘New’ order types time series shows that there was a steep drop in orders near the 15th May 2021. (See images below)

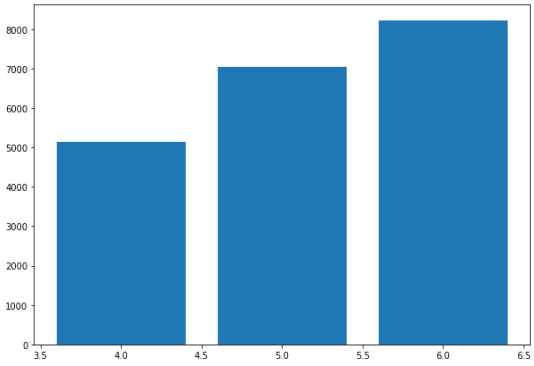



Order_type = ‘New’ (Barplot)

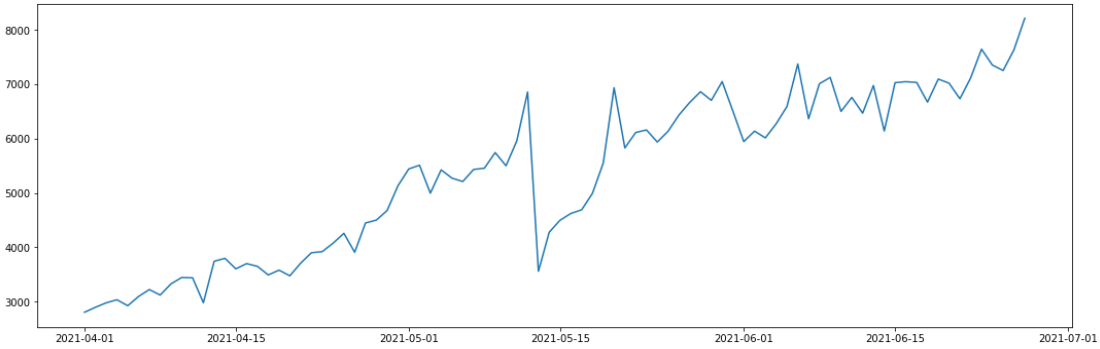



Order_type = ‘New’ time series

3. While the ‘Repeat’ order types saw a steady increase from April 2021 - June 2021.




Order_type = Repeat’ (Barplot)




Order_type = ‘Repeat’ time series

4. **A point to note is that**: The difference between the ‘New’ and ‘Repeat’ order_type is significant. The **‘Repeat’** orders start from around 3000 orders and have a range of about 8000 orders while the **‘New’** orders have a range from (approx. 800-2400) orders.

**Answer2:**

One of the best ways to find a drop in sales is to check for customer complaints. The analysis can be found under **‘Airlift_Technologies_dataset4.ipynb’ (complaints).**

While there are a number of complaints a customer could have, analysis showed that the major Reasons were ‘Product Quality/Damage’ followed by ‘Delivery Issues’. Most of the **complaints** were in May 2021. The ‘Delivery Issues’ can be further linked to Commute time problems and ‘Warehouse issues too’: **The average time a product was out of stock was 9.74 hours**. The most amount of **sub-complaints** were: ‘Rider Was Later’,’Damaged’. Please see the figures below:

## Most frequent Complaint Reason

```
df4.complaint_reason_title.unique()
```

```
1]: array(['Payment', 'Technical', nan, 'Product Quality/Damage', 'Delivery',
       'default'], dtype=object)
```

```
df4.complaint_reason_title.nunique()
```

```
2]: 5
```

```
df4.complaint_reason_title.value_counts()
```

```
5]: Product Quality/Damage    36334
    Delivery                  21813
    Payment                   11389
    Technical                 11026
    default                     305
    Name: complaint_reason_title, dtype: int64
```

## Months with the most Complaints

```
df4.month.value_counts().plot(kind='barh')
```

```
39]: <AxesSubplot:>
```