

ADV Help Mailbox Analysis

Fall 2020

Author: Ethan Creagar

Contents

0.0.1	Description:	1
0.1	Summary of Findings:	2
0.1.1	Inbox basics:	2
0.1.2	Time Trends:	2
0.1.3	Category Trends:	2
0.1.4	Completed Ticket Trends:	3
1	Part 1 - Inbox Basics	3
1.0.1	a: How many emails have we received?	3
1.0.2	b: On average, how many tickets do we solve?	3
1.0.3	c: How have these trends changed over time?	4
2	Part 2: Time Trends	4
2.0.1	What are our times of highest volume?	4
3	Part 3: Category Trends	10
3.1	a: What is the category breakdown on incoming tickets?	10
3.2	b: What does this breakdown look like by time period?	12
4	Part 4: Completed Ticket Trends	22
4.1	a: What is the breakdown of tickets completed by category?	22
4.2	b: What is the breakdown of completed tickets by technician?	23
5	Discussion of Limitations and Future Considerations:	25
5.0.1	Limitations:	25
5.0.2	Future Considerations:	26

0.0.1 Description:

This project, authorized by Nate Williams and Colorado State University Advancement, explored the history of the ADV Help Outlook inbox from 2016-2020 to search for insight on trends and provide baseline numbers for the helpdesk. Specifically, the following points will be addressed:

-
- 1) **Inbox basics:** How many tickets has the helpdesk received? On average, how many of these tickets do we solve? How has this changed over time?
 - 2) **Time Trends:** What are our times of highest ticket volume? How can we utilize staff and resources best to fit with these times?
 - 3) **Category Trends:** What is the category breakdown on incoming tickets? What is does this breakdown look like by time period?

- 4) **Completed Ticket Trends:** What is the breakdown of tickets completed by category? What does this breakdown look like by technician?
-

These questions give us insight into the workings of the helpdesk and help us set goals and be better prepared for the issues of Advancement employees in the future.

0.1 Summary of Findings:

The insights we can pull from the inbox are stated below.

0.1.1 Inbox basics:

How many tickets has the helpdesk received? On average, how many of these tickets do we solve? How has this changed over time?

- **ADV Help has received 15,377 emails over the last 5 years.** This is a trend that has been growing - on average we receive about 5 more emails every month than the last. We solve about 95 percent of tickets ($\pm 5\%$).

0.1.2 Time Trends:

What are our times of highest ticket volume? How can we utilize staff and resources best to fit with these times?

- **Afternoons** are typically busier than mornings. We have received almost 53% of our tickets in the afternoon, although that trend has changed this calendar year. In particular, we see our most emails between 9:00 and 11:00 AM and 1:00 and 3:00 PM
- **Tuesdays and Wednesdays** are our busiest days of the week. Monday afternoons are busy, while mornings have much less volume. Friday is our least busy time, both morning and afternoon.
- Generally, **October, August and February** are times of more emails, while **December** is our least-emailed month.

0.1.3 Category Trends:

What is the category breakdown on incoming tickets? What does this breakdown look like by time period?

- **About 20 percent of the total tickets we receive are Hardware tickets, our highest category.** Email and Network tickets are next, with just over 8 percent each, then Software tickets.
- **As opposed to our total tickets, we receive more Remote tickets in the morning.** Other categories do not see this same trend.
- **Software tickets are growing at a faster rate than total tickets are.** We see that year by year, a higher percentage of our total tickets are software based.

- We tend to see a **spike in personnel tickets in August**. This has been a consistent trend across the last few years.
- **We see the most email tickets in January**. This is a significant finding because January is generally one of our lowest volume times for lowest tickets. We may want to examine whether there is a reason this would be the case.
- **Purchasing tickets see a spike in September**. We should be on the lookout for more tickets involving purchasing, hardware replacements, etc. at this time of the year.
- **We see the most network tickets on Tuesday mornings**. It is not out of the ordinary to see an increase in tickets on Tuesdays, but we know that mornings are generally less busy than afternoons. We also don't see a consistency in increased Network tickets in the mornings, which could lead us to believe that Tuesday morning Network issues may need to be examined further. This trend is most pronounced in 2020.

0.1.4 Completed Ticket Trends:

What is the breakdown of tickets completed by category? What does this breakdown look like by technician?

- **We generally complete tickets at a similar rate to which we receive them in each category**. The Hardware, Remote, and Software ticket categories have the biggest discrepancies in them between received and completed.
- **New hires do not have much of a difference in tickets completed by category than more experienced employees do**. We do see that (with a small sample size) new technicians complete a higher percentage of hardware and software tickets, whereas more experienced technicians complete more network tickets.

1 Part 1 - Inbox Basics

In Part 1, we look to answer some basic questions about the inbox, such as the quantity of tickets received and how this has changed over time.

1.0.1 a: How many emails have we received?

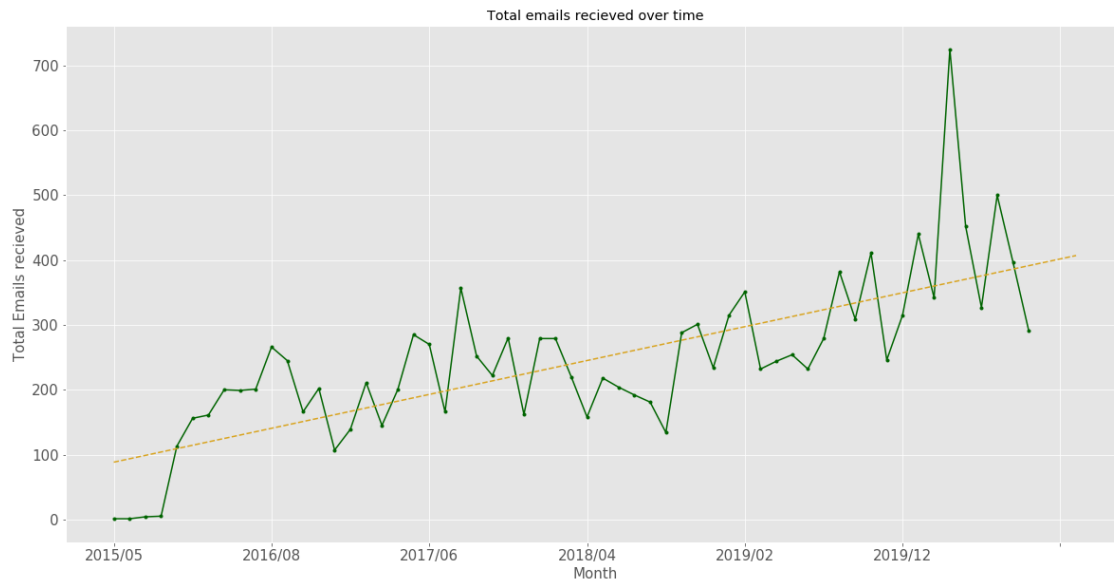
Through the last 5 years, ADV Help has received 15,377 emails.

1.0.2 b: On average, how many tickets do we solve?

Completed 5660 out of 6008 unique tickets: 94.21% (plus/minus 5%)

By indexing the data to only include the tickets whose categories contain the word "Done", we can see how many tickets we have completed. We can compare this against only unique emails, defined as emails without "RE:" at the beginning of them as this would indicate that the email was part of a chain, and find the percentage of tickets we've completed. This percentage relies on accurate ticketing of data, so I built in a 5% reduction and Confidence Interval of 5% to adjust for marking the same ticket as "Done" twice at different points in its chain. $\pm 5\%$ was chosen by exploring a subset of tickets and examining how many were marked done at two different times in the inbox.

1.0.3 c: How have these trends changed over time?



Here we see the trend of our emails received from February of 2016 to April of 2020. We see that there is an upward trend in this data and we can calculate this trend below.

Intercept: 88.303

Beta 1: 5.226

The line of best fit through the data can be quantified as $\hat{y} = 88.303 + 5.226\hat{x}$. β_1 is likely affected by the outlier of March 2020.

Over time, the number of emails received by ADV Help has grown at a rate of about 5.23 emails per month, or about 63 emails per year.

2 Part 2: Time Trends

In part 2, we want to examine how many tickets we get in different time periods, whether this be morning vs. afternoon, months, years, or a combination of these things.

This can help us understand inbox trends in order to allocate staff and resources better.

2.0.1 What are our times of highest volume?

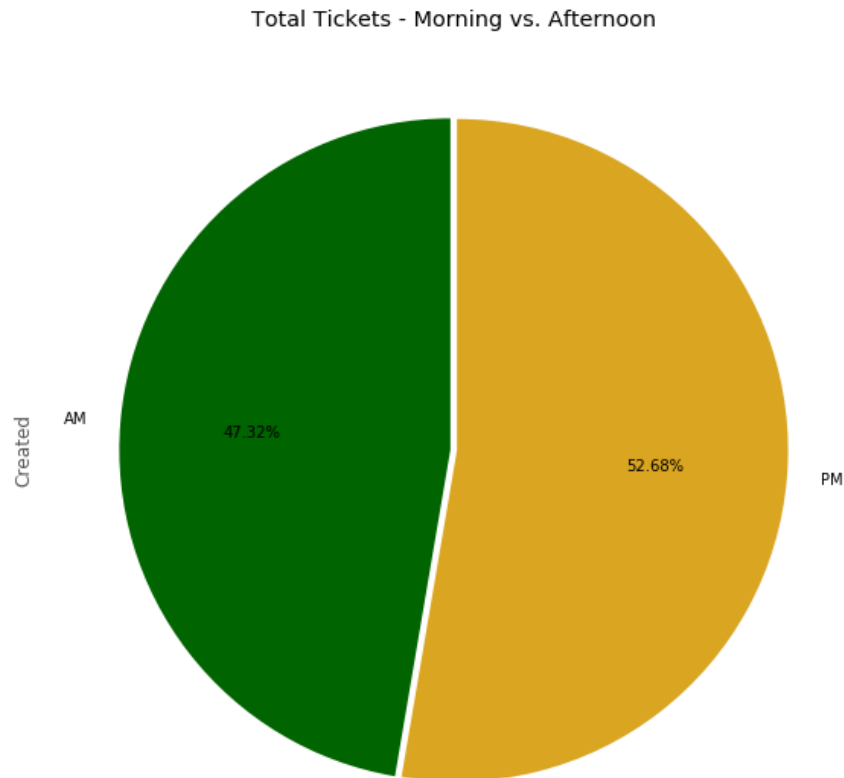
Methodology

We can plot emails received by time frame with many options, including:

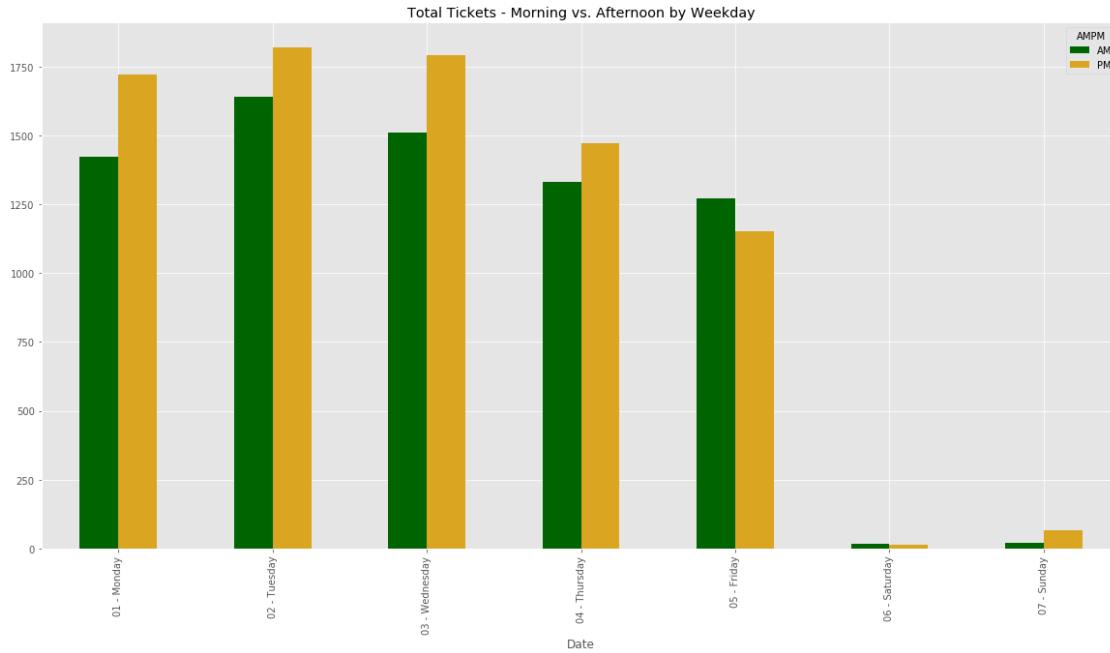
- “Date”: Which weekday - Monday to Sunday - the email was received
- “AMPM”: If the email was received in the morning or afternoon
- “Hour”: Which hour of the day - 1 to 23 - the email was received
- “Minute”: Which minute - 1 to 60 - of the hour the email was received

- "Month": Which month - 1 to 12 - the email was received
- "Day": Which day - 1 to 31 - the email was received
- "Year": Which year the email was received in

With this information, we can plot something simple like whether we receive more tickets in the morning or afternoon:

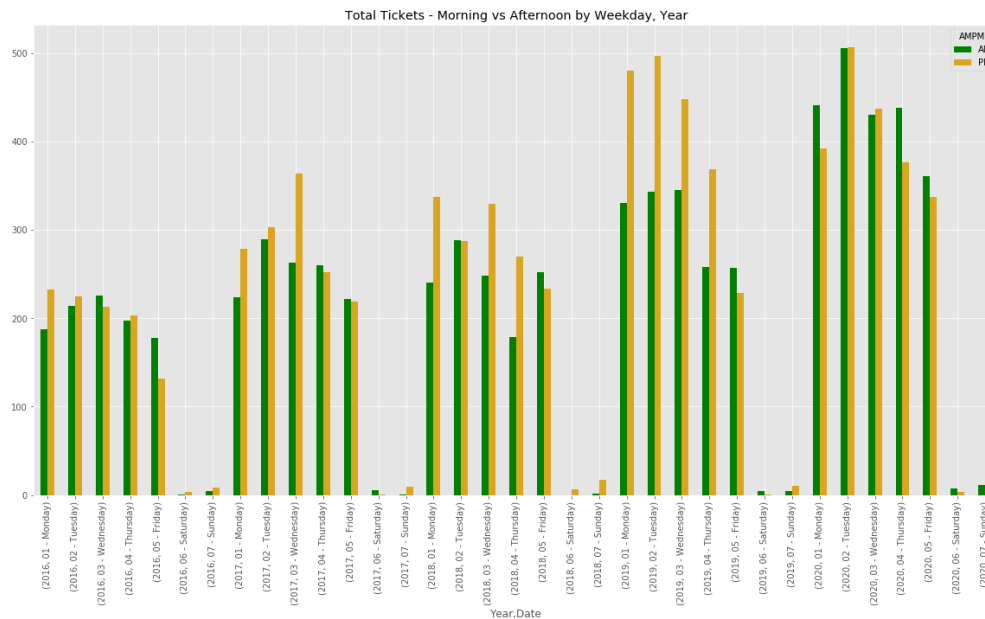


We see that in our history, we have received more emails after 12:00 PM than before. This could be helpful, but a more in depth version might include weekday in it as well to be more thorough. Including weekday, for example, would look like this:



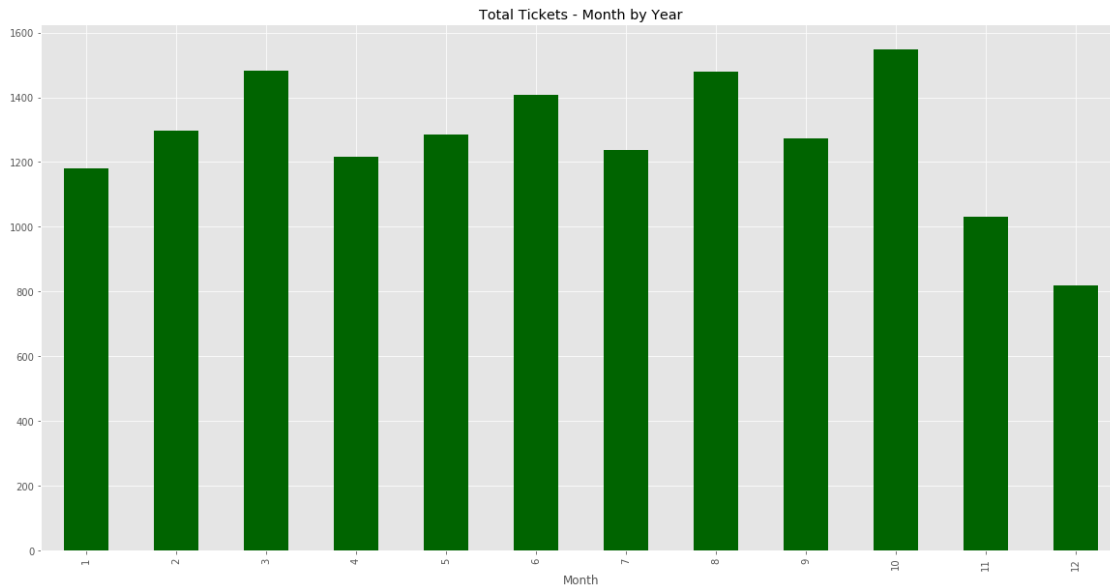
Here we can see that amount of tickets received in the afternoon is higher earlier in the week and trends downward as the week goes on. Tuesdays and Wednesdays are our busiest days, while Fridays see the least emails of any weekday. Friday mornings also tend to be busier than Friday afternoons, the only day where this is true. Mondays, however, show the largest difference between afternoon and morning.

An even more in depth version might ask whether this trend has changed over the last few years. Including year in the plot would look like this:

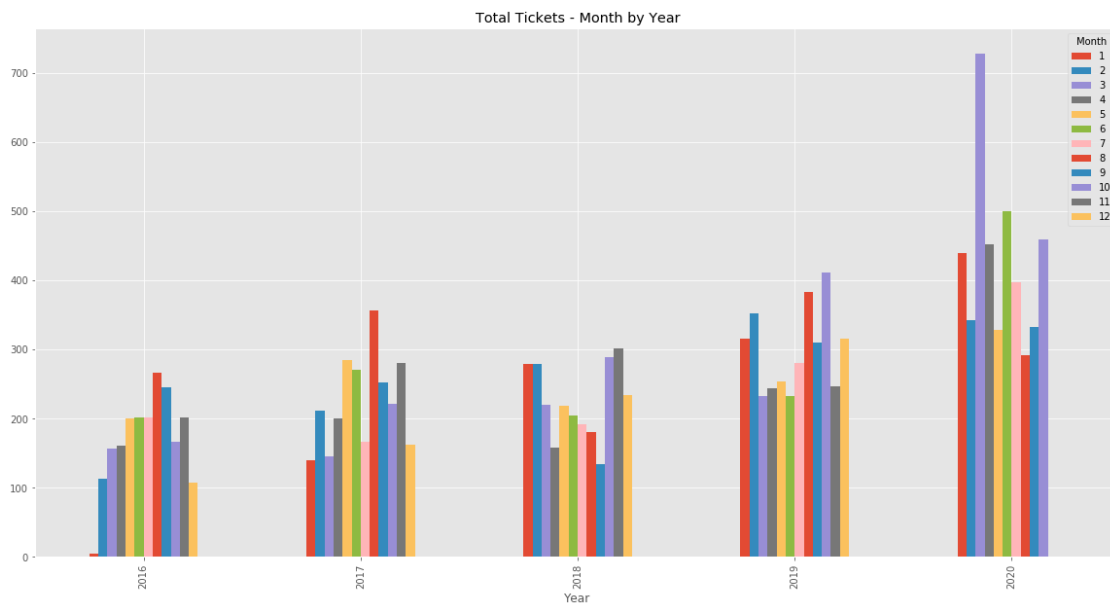


Although in the past we've received more emails in the afternoon, this trend is not true for 2020. So far, we are receiving more emails in the morning this year.

There are many interesting combinations that can give us trends of when we should be prepared for more emails.



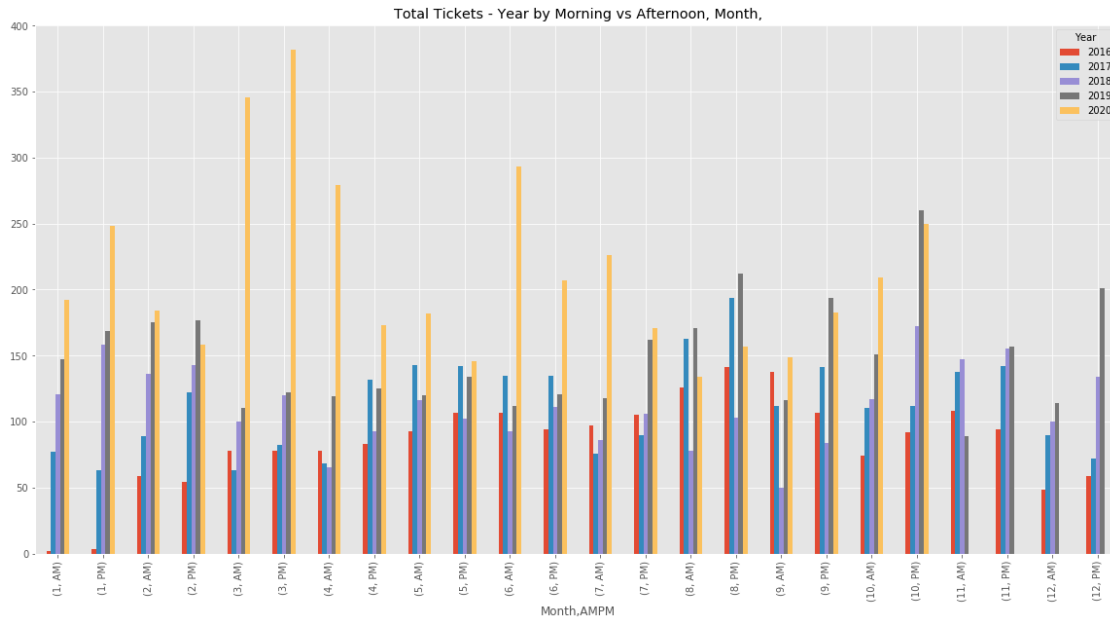
Here, we see that March is one of our largest month for emails received. However, this is misleading, as we see in the plot below.



By organizing the variables this way, we can see trends by year colorized by month. This allows us to see if there are any trends with certain months being more or less busy. If we had just broken this down by month, the data shows March as one of our busiest months, but looking at it this way we can see that, apart from the outlier in 2020, the opposite is usually true and March is not generally a large month for us.

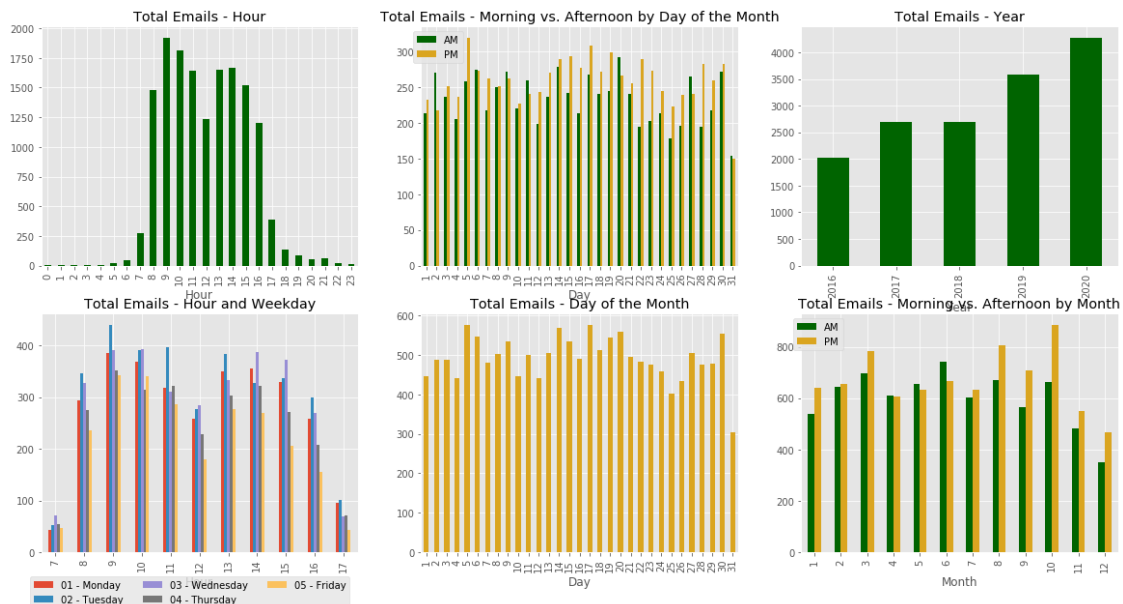


This plot shows us the inverse of the last plot - each month, colorized by which year the ticket was received in. This gives us insight to be able to compare the first few months of a year to previous years or specific months against previous years. We can see here that most months seem to be rising over time, which agrees with our previous insight that we are receiving consistently more emails over time. We can also see that even before lockdown, January and February were shaping up to be as busy or busier than last year, meaning not all of the influx of emails should be attributed to working from home and that we were likely to continue to see growth in the inbox regardless.



Here we break the last two down further by splitting the groups into Year, morning and afternoon. We can use this plot to see whether changes over time are consistent between morning and afternoon. We can see that in some months, like October, we have been historically more likely to receive emails in the afternoon.

More examples of plots with insights are shown below.



- **Total Emails - Hour:** We receive the most emails between 9 and 10 o'clock, followed by

between 10 and 11 o'clock and between 2 and 3 o'clock.

- **Total Emails - Hour and Weekday:** The amount of emails we receive by hour is fairly constant across the day of the week. We see more emails between 9:00 and 12:00 on Tuesdays than any other day, while we see the most emails between 2:00 and 4:00 on Wednesdays.
- **Total Emails - Morning vs. Afternoon by Day of the Month:** We see a similar breakdown to the plot below here, but we can see that we get more emails in the afternoon later in the month than we do earlier in the month.
- **Total Emails - Day of the Month:** We see the most emails on the 5th and 6th of the month, followed by the 30th. We receive slightly less emails later in the month than earlier in the month, but it's not a very significant difference.
- **Total Emails - Year:** We have looked at this in different ways, but this plot shows us that this year will be our highest emailed year by far.
- **Total Emails = Morning vs. Afternoon by Month:** we see more email in the afternoon in the later months than we do in the first 6 months.

There are many more combinations of plots we could create, but the insight drawn from these is valuable.

To answer our question, we can summarize some **trends of the times of increased and decreased volume**:

- **Afternoons** are typically busier than mornings. We have received almost 53% of our tickets in the afternoon, although that trend has changed this calendar year. In particular, we see our most emails between 9:00 and 11:00 AM and 1:00 and 3:00 PM
- **Tuesdays and Wednesdays** are our busiest days of the week. Monday afternoons are busy, while mornings have much less volume. Friday is our least busy time, both morning and afternoon.
- **Earlier in the Month** tends to be slightly busier than later in the Month.
- **The Month of the year** in which we received the most email does not have much of a constant trend. Generally, **October, August and February** are times of more emails, while **December** is our least-emailed month.

3 Part 3: Category Trends

In this section, we break our inbox down into categories to see if we can find any trends based on the different categories. We look to explore this by different time period like we did in part 2.

3.1 a: What is the category breakdown on incoming tickets?

Using a function to sort emails, we can go through each subject line and find keywords that might mean a ticket is likely to be in a certain category. Some examples of these keywords are as follows:

Email Tickets: "outlook", "email", "mailbox", 'delegate', 'calendar', 'spam'

Remote Tickets: "remote", "rdp", "remote desktop", "pulse", "duo", "vpn"

Hardware: "monitor", "webcam", "keyboard", "dock", "mouse", "cables", "printer"

Network: “Network drives”, “internet”, “connection”, “domain”, “wifi”, “distribution list”

Adobe: “adobe”, “acrobat”, “pdf”, “creative cloud”, “photoshop”

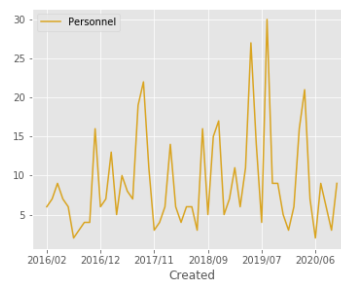
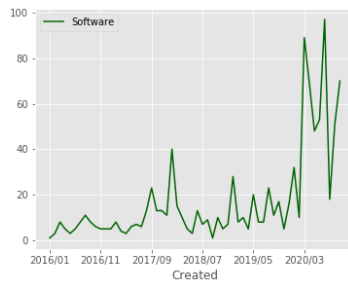
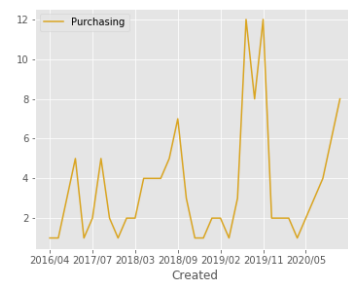
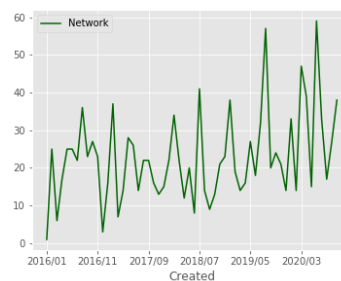
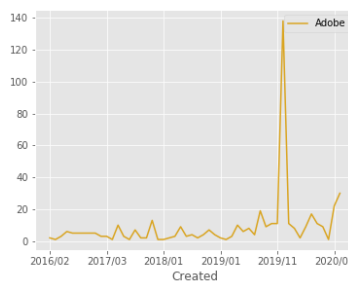
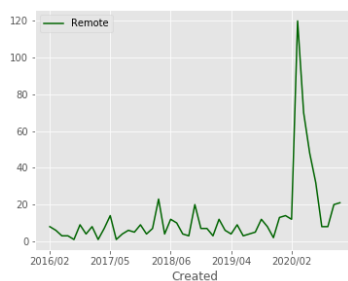
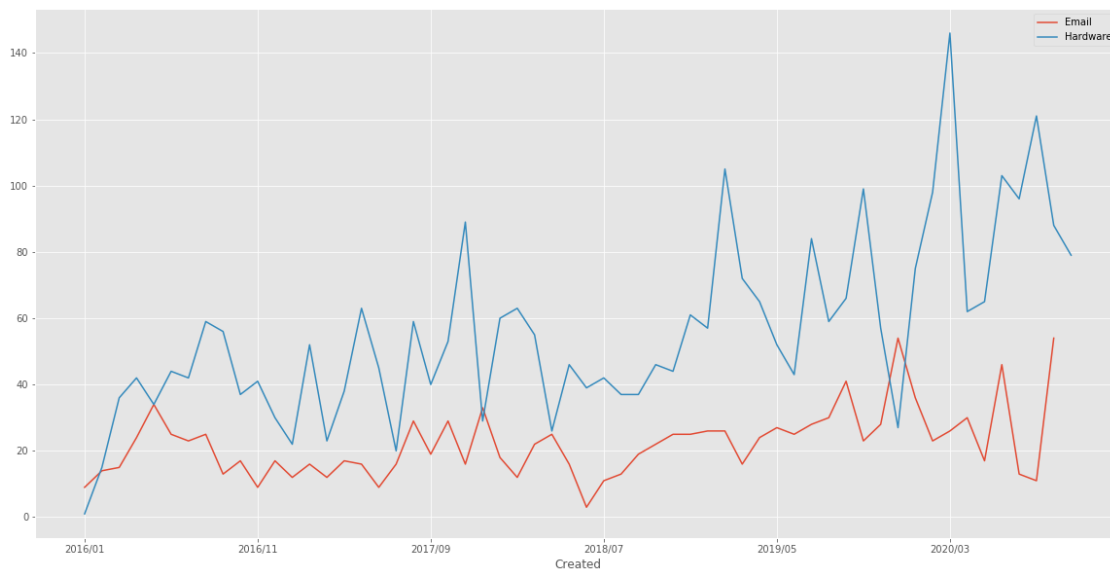
Software: “install”, “download”, “activation key”, “software”, “licensing”, “word”, “excel”

Purchasing: “order”, “delivery”, “kuali”, “cdw-g”, “pricing”

Personnel: “employee”, “hire”, “on-boarding”, “resignation”, “departures”

Phone: “call”, “missed call”, “voicemail”, “phone”

We can plot Categories by time or against one another by time. For example, below is a plot of Email category tickets vs Hardware category tickets over the last 4 years.



The rest of the categories are plotted above. We can see that most of them have a spike recently and are generally trending upward. Most of these spikes are easily explainable - for instance, the **Remote** category spikes around March of 2020, when the division was starting to work from home and needed our remote connection guides and software. **Adobe** has a spike in November/December of 2019, the month when the technicians were upgrading all machines to the new Adobe Acrobat 2017.

The breakdown of total emails by subject can be seen below.

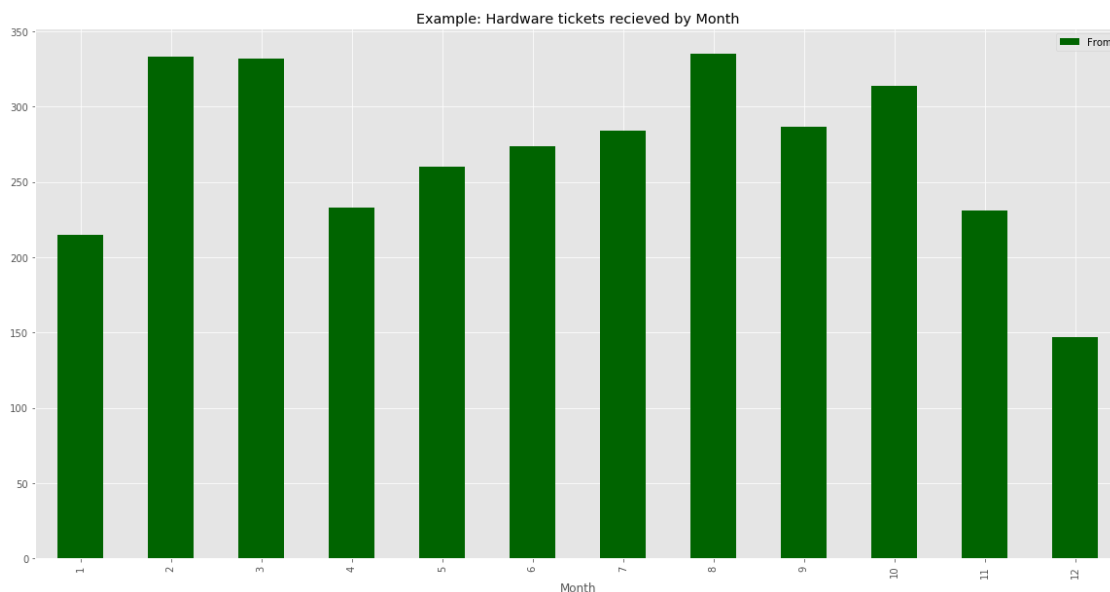
[36] :

	From	Total	Percentage
Other	5584	15251	36.6
Adobe	464	15251	3.0
Email	1264	15251	8.3
Hardware	3245	15251	21.3
Network	1324	15251	8.7
Personnel	497	15251	3.3
Phone	1124	15251	7.4
Purchasing	124	15251	0.8
Remote	624	15251	4.1
Software	1001	15251	6.6

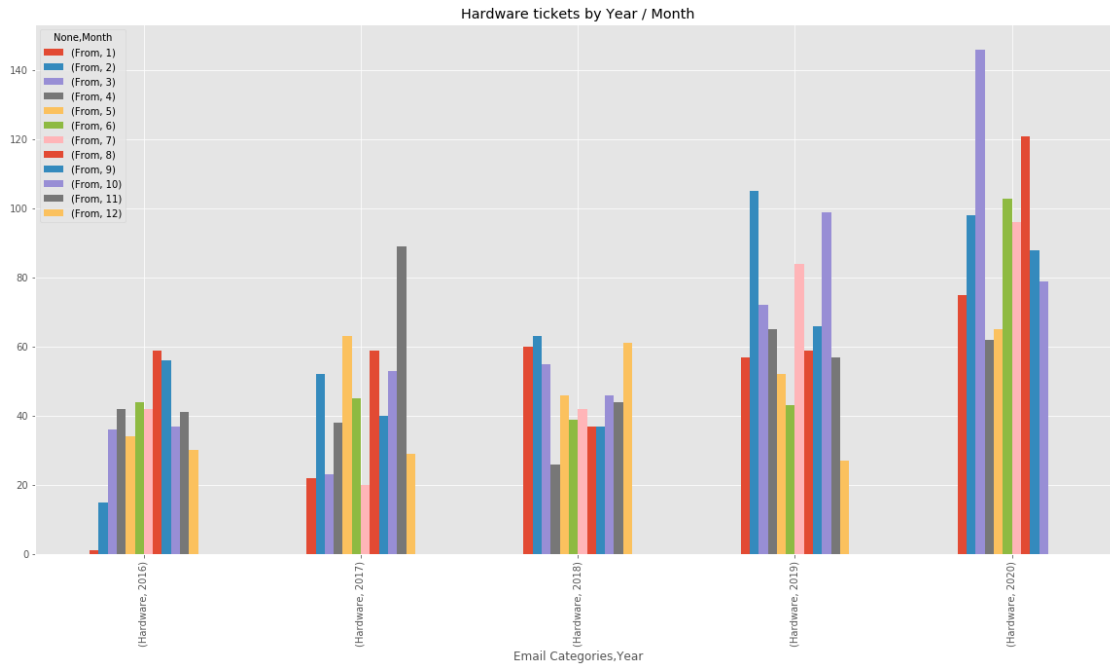
3.2 b: What does this breakdown look like by time period?

We might want to know if there are any trends of categorized tickets to know if we see more errors or slowness at any times in specific.

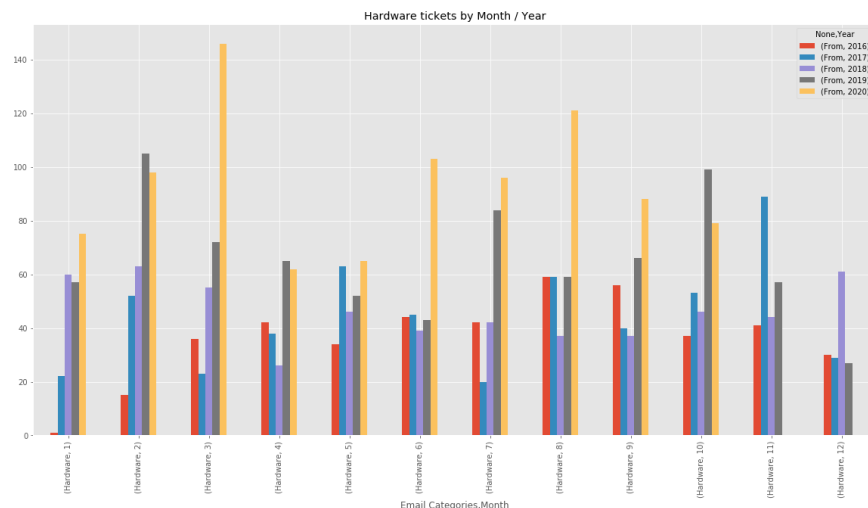
Our Group Categories and Plot Categories functions can give us the same insight as we received above, but this time broken down into categories so we can see when we are receiving the most of a certain type of category in a more detailed way than just over each month like above.



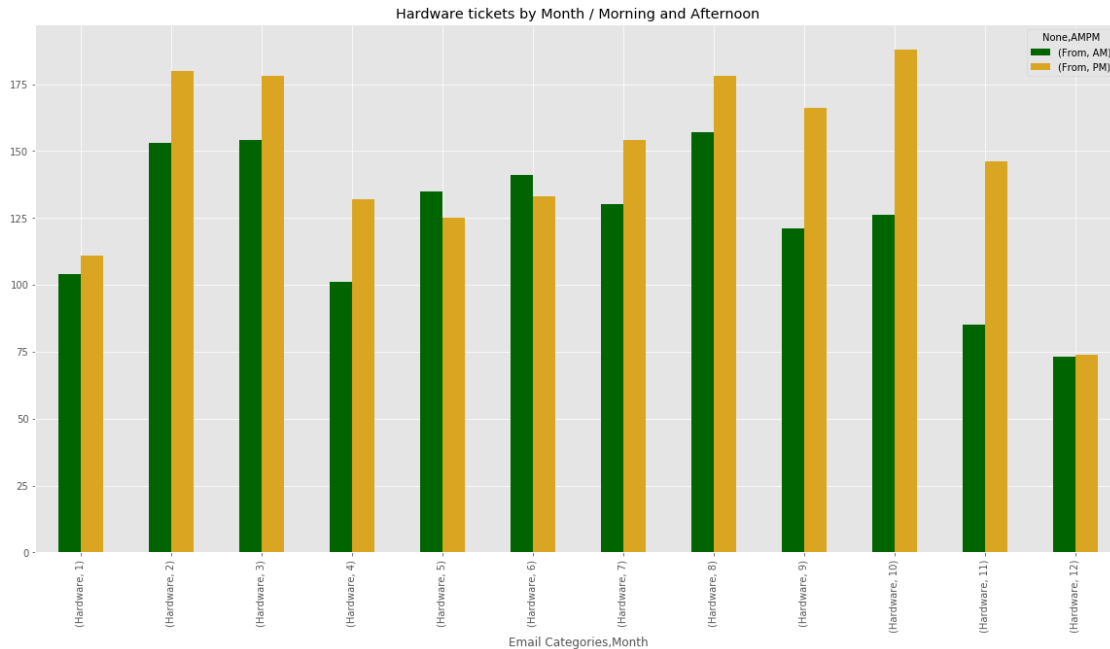
For example, here we see the **Hardware** category broken down by month. Again, this may not be as insightful as we'd like and we might have a bit of a bias toward March and April due to the large spike of emails in that time period of 2020. We can break this down further to get a better idea of this category's time distribution.



We see here that we are correct - There was a big spike in February and March of 2020. We could, however, argue whether the February spike was out of the ordinary based on the fact that we seem to see many **Hardware** tickets in February each year. We also seem to see spikes in **October/November**.

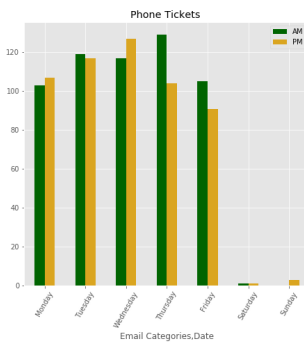
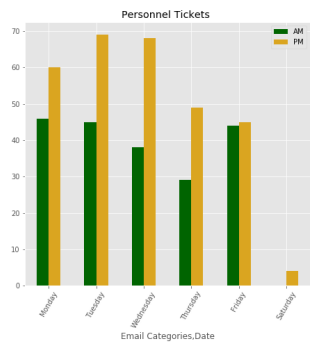
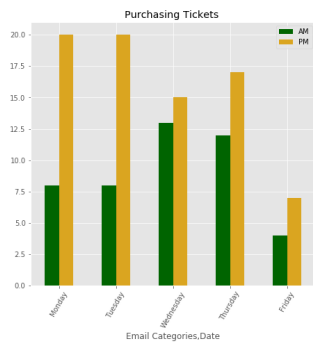
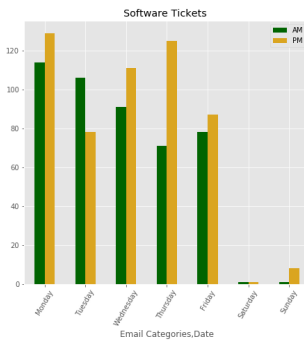
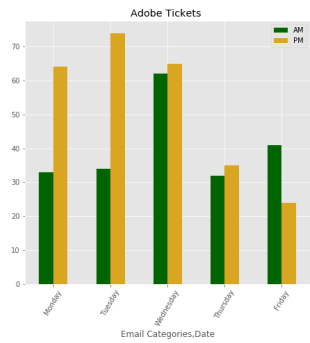
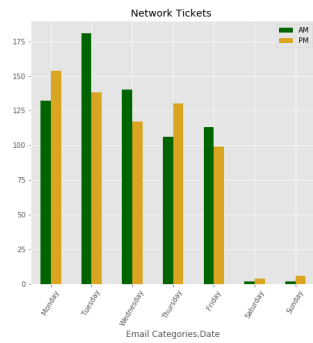
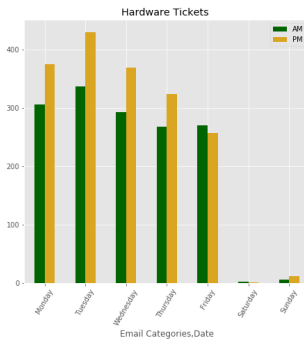
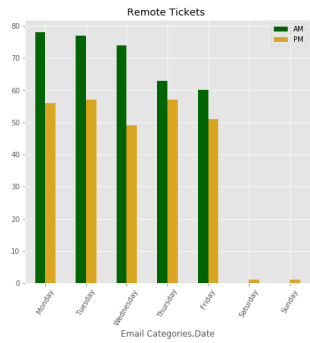
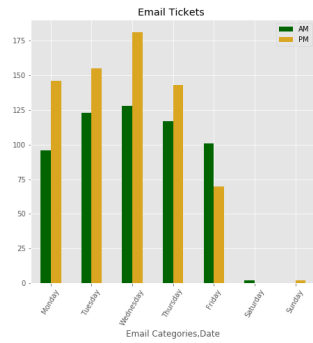
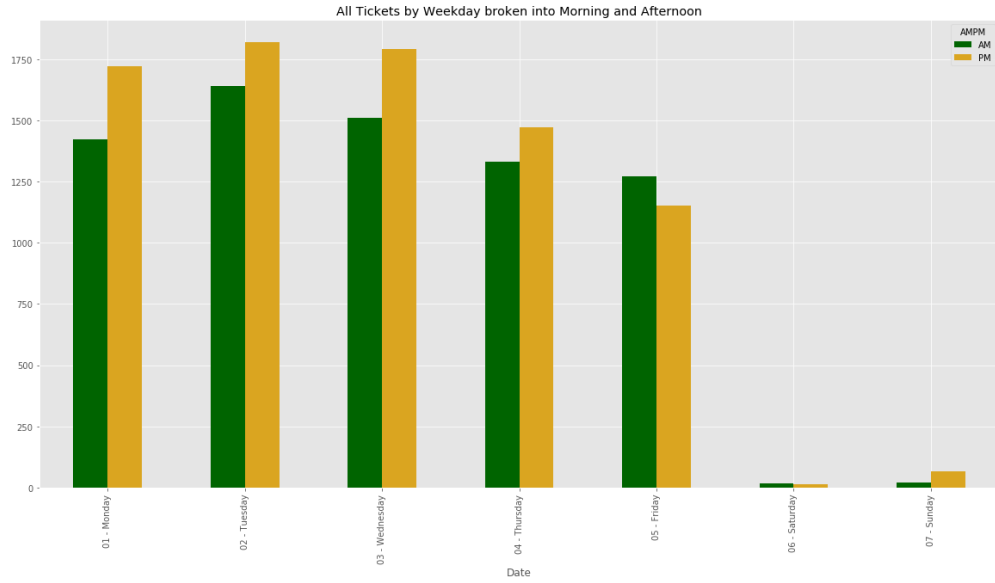


The first plot breaks down each year by month so we can see trends by month inside each year, while the second breaks down each month by year to see how our number of emails received in each month have grown or shrunk by year. This gives us the insight that **like emails in general and most categories of emails, Hardware tickets are rising fairly consistently by year.**



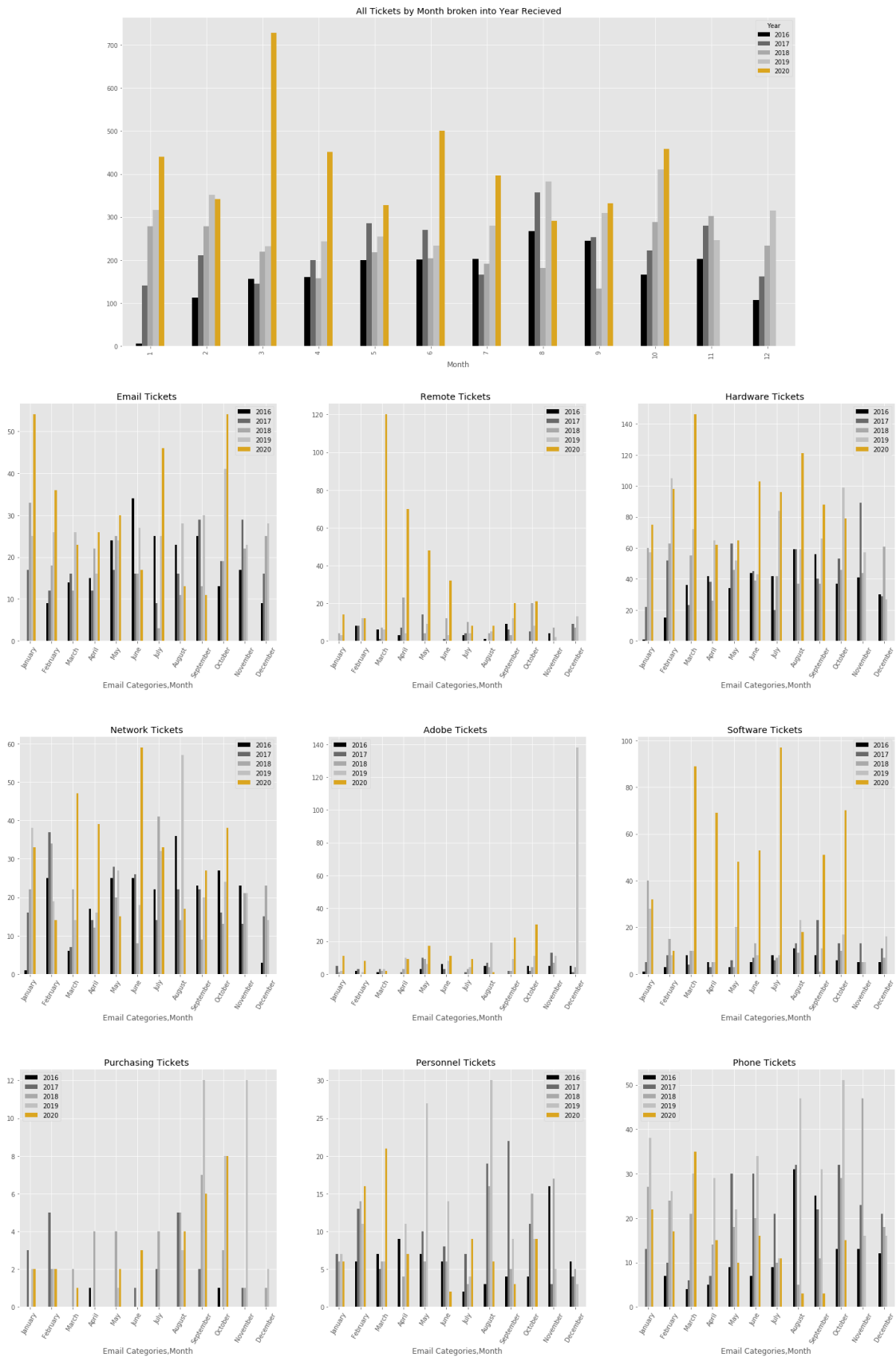
This plot tells us that **certain months, such as October and November, and have generally had a larger discrepancy between the morning and evening emails.**

Now that we understand the basics of examining categorized tickets by time frame, we can plot each category against our overall inbox. By doing this, we can see any differences that stand out in any of the categories. We can break these time frames down by time periods that might be of interest, like Year, Month, Weekday, and AM/PM.



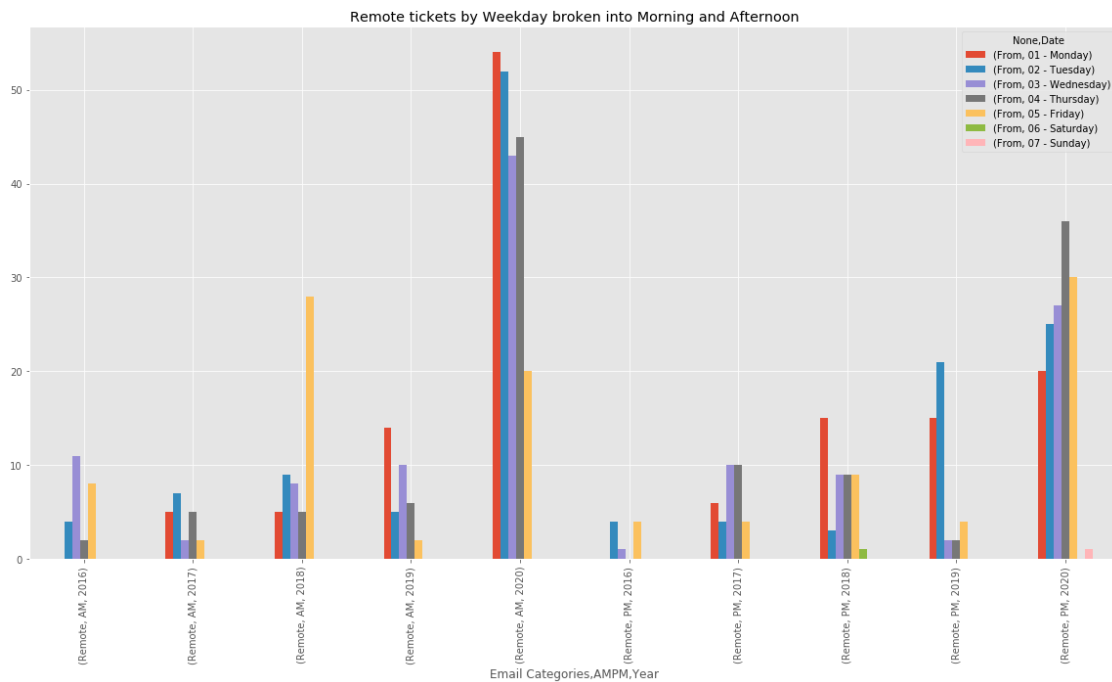
In this plot, we can see that most categories seem to follow the same trend as our overall inbox of receiving more emails in the afternoon. However, Remote tickets (tickets containing keywords such as "Pulse", "VPN", "Duo", and "Remote Desktop" are received more often in the mornings. This can give us the insight that **we should expect more tickets about remote connection in the mornings.**

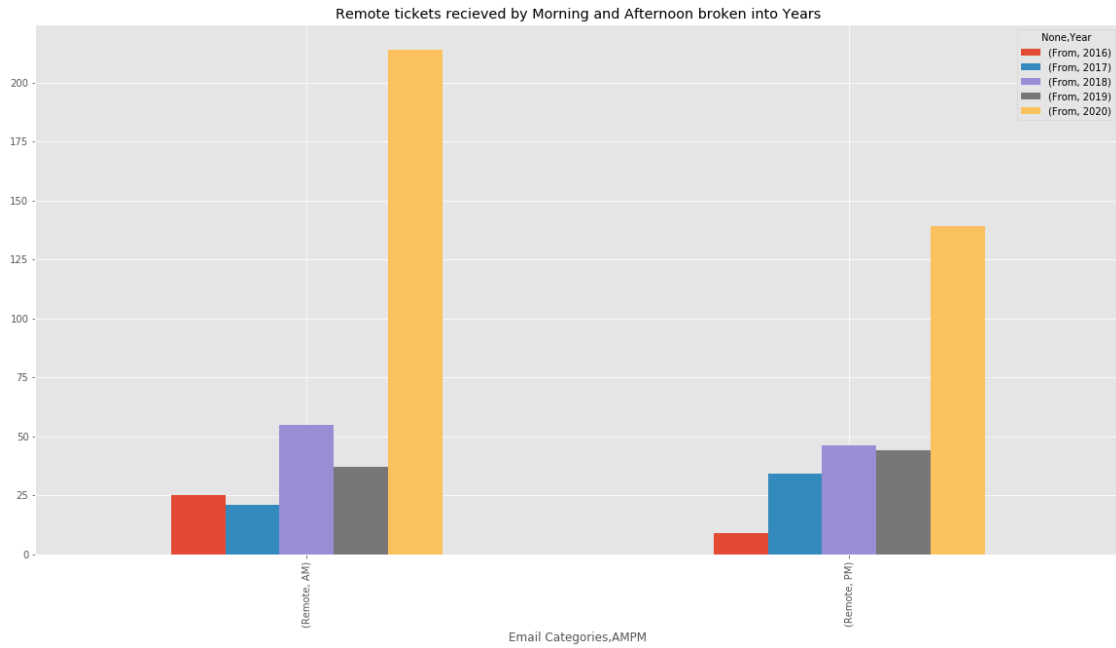
We can create similar plots for Year and Month Trends:



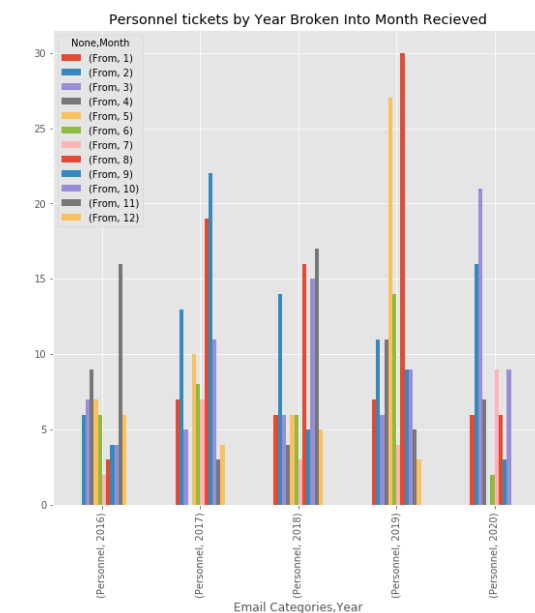
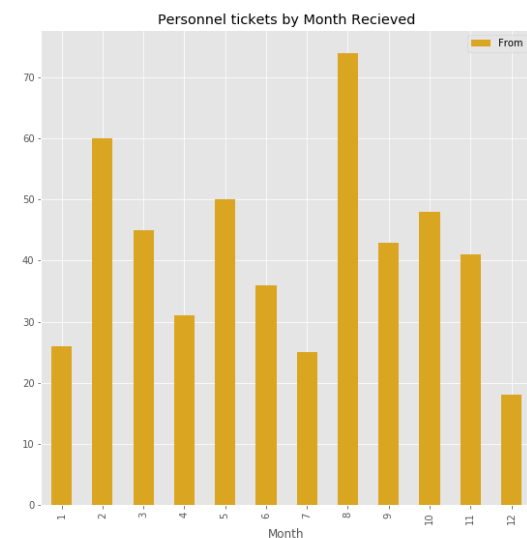
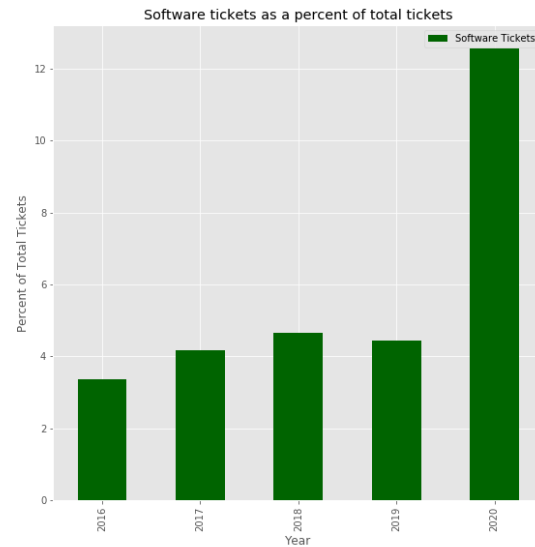
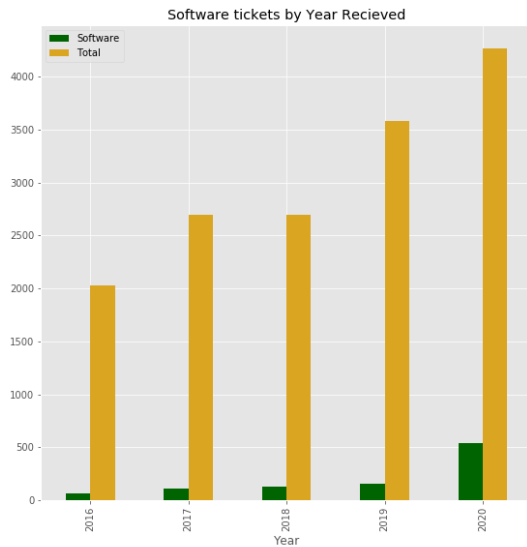
These plots contain much more information. We can see that our general trends see August - October/November as the high points of the year. This is also true for many of our categories, but some differences that stand out between the plots include: **Phone tickets do not seem to be rising at the same rate as other tickets. Software (including keywords “download”, “install”, “license”, etc.) tickets have seen the biggest rise of any category this year.**

We’ve seen that we get a lot more Remote tickets in the mornings than we do in the afternoons, which could help us be more on the lookout for these tickets in the mornings. We can see that this trend is also heavily influenced by 2020.

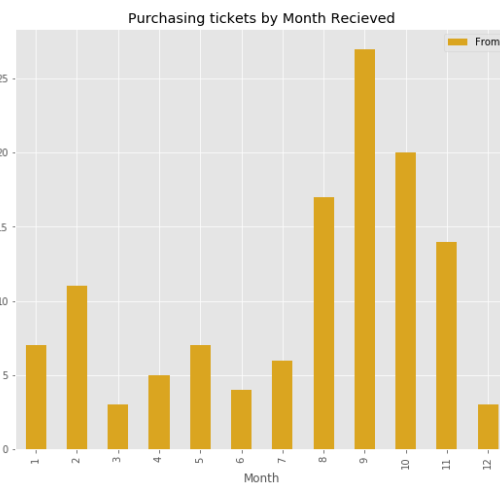
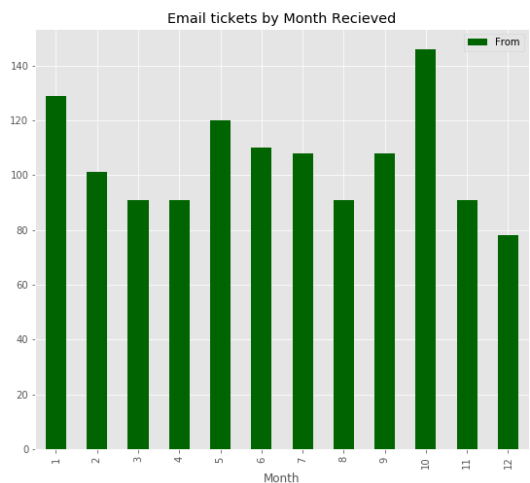
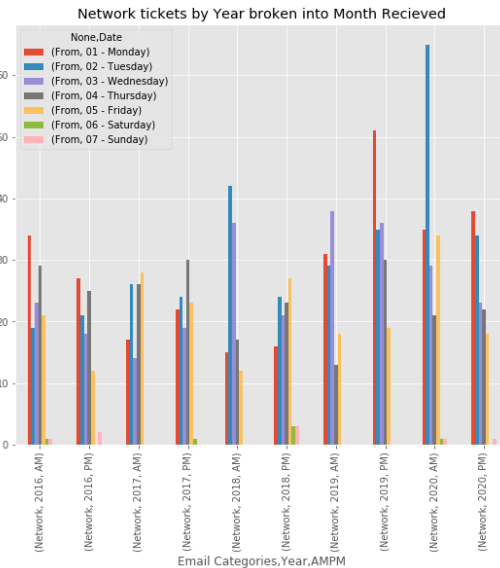
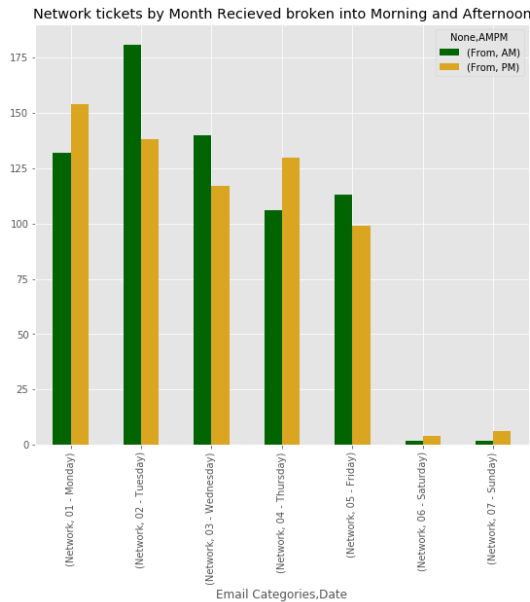




Examining many more plots and searching for interesting and/or unexpected trends yielded the insights explored in the plots below.



- **Software Tickets:** Plotting software tickets vs total tickets by year received, we can see that **software tickets**, tickets including keywords such as “download”, “update”, “software”, “installation”, and “licensing”, **seem to be growing at a faster rate than our tickets overall**. As a total percentage of our tickets, software tickets are also growing year-by-year with a spike in 2020.
- **Personnel tickets:** We see that we get a **spike in personnel tickets**, tickets including keywords such as “Hire”, “On-boarding”, “Welcome”, and “Resignation”, **in August**. Looking at the plot of Personnel tickets by year broken into month received, we can see that **this spike in August/September is fairly consistent over at least the last three years**, so this spike is fairly reliable.



- **We see the most network tickets on Tuesday mornings.** It is not out of the ordinary to see an increase in tickets on Tuesdays, but we know that mornings are generally less busy than afternoons.
- **Email tickets by month received:** By breaking down our “email” tickets, tickets with keywords like “Outlook”, “Mailbox”, and “Email”, we can see that we see our largest spike in January. This may be an interesting trend to look into as it counters our findings above that January is generally not a time of large email volume.
- **Purchasing tickets by month received:** Our purchasing tickets, tickets with keywords such as “order”, “purchase”, “replacement”, or “delivery”, sees by far its largest spike in September.

4 Part 4: Completed Ticket Trends

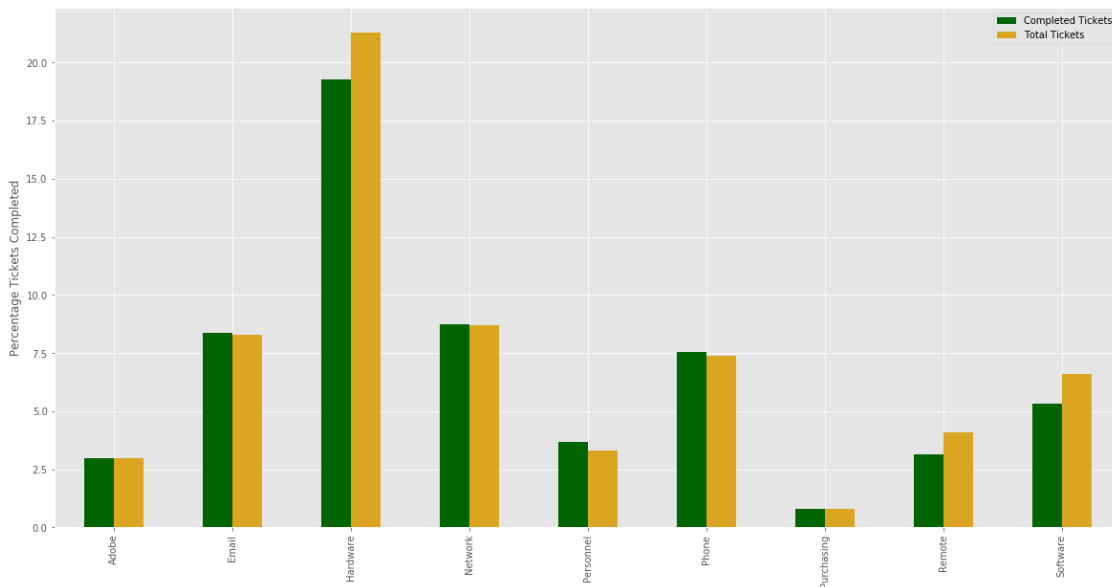
In part 4, we want to answer questions about the breakdown of our total tickets in each category. We can use this information to inform decisions about what tickets to devote the most resources to, how we should train employees, and more.

4.1 a: What is the breakdown of tickets completed by category?

We now have a breakdown of what tickets we receive the most of and when, which can be useful to know what kind of tickets we should devote the most resources to. It might also be interesting to see what kind of tickets we complete at the highest rate to see if we struggle in some categories more than others. Note that these percentages lean on the side of too high due to tickets marked Done twice.

	Percent - Completed	Percent - Total
Adobe	2.97	3.0
Email	8.37	8.3
Hardware	19.26	21.3
Network	8.73	8.7
Personnel	3.69	3.3
Phone	7.56	7.4
Purchasing	0.81	0.8
Remote	3.15	4.1
Software	5.31	6.6

We add some noise to the Completed tickets group to account for the lower percentage of Uncategorized tickets and plot the two data frames side-by-side.



We can see that our numbers follow pretty closely, and with the inaccuracy of the done tickets we

can say that most of the percentages are likely very close to each other without duplicates. The interesting columns then are the ones that have lower or very close percentages in the Done data frame. These are **Hardware, Remote, and Software**. If our data is accurate, this would mean that **these are the tickets that go uncompleted most often**. This could be something to look out for in the future.

4.2 b: What is the breakdown of completed tickets by technician?

Dividing the tickets by technician, we can compare what kind of tickets a technician responds to most vs what kind of tickets we receive the most to see if our technicians have any obvious gaps in the type of tickets they respond to. For instance, here are the number of tickets I've completed (those marked as Done - Ethan) vs the total numbers.

	Percents - Ethan	Percents - Total
Adobe	7.8	3.0
Email	9.5	8.3
Hardware	23.1	21.3
Network	7.4	8.7
Personnel	3.7	3.3
Phone	5.6	7.4
Purchasing	1.7	0.8
Remote	6.9	4.1
Software	5.7	6.6

We can see that my responses (on top) have a bit of an advantage in the Adobe and Email categories, which means I respond to tickets of those sort at a higher rate than our historical average. However, I respond at a lower percentage to Network, Phone, and Software category tickets, meaning I may have a bit of work to do in terms of finishing those tickets.

We can gather this data for any technician whose tickets are still marked "Done - Technician" in our inbox. For example, here we can see Nate's tickets:

	Percent - Nate	Percent - Total
Adobe	2.4	3.0
Email	9.0	8.3
Hardware	19.5	21.3
Network	10.7	8.7
Personnel	5.4	3.3
Phone	10.2	7.4
Purchasing	0.5	0.8
Remote	2.8	4.1
Software	5.6	6.6

Comparing the emails from Done - Nate to the total ticket breakdown, we can see that his responses follow the distribution pretty closely. This would make sense, since he's solved about 1/6 of the total ADV Help tickets. Nate's responses also lean toward Network tickets slightly, which we would expect with what kind of tickets he tends to take care of.

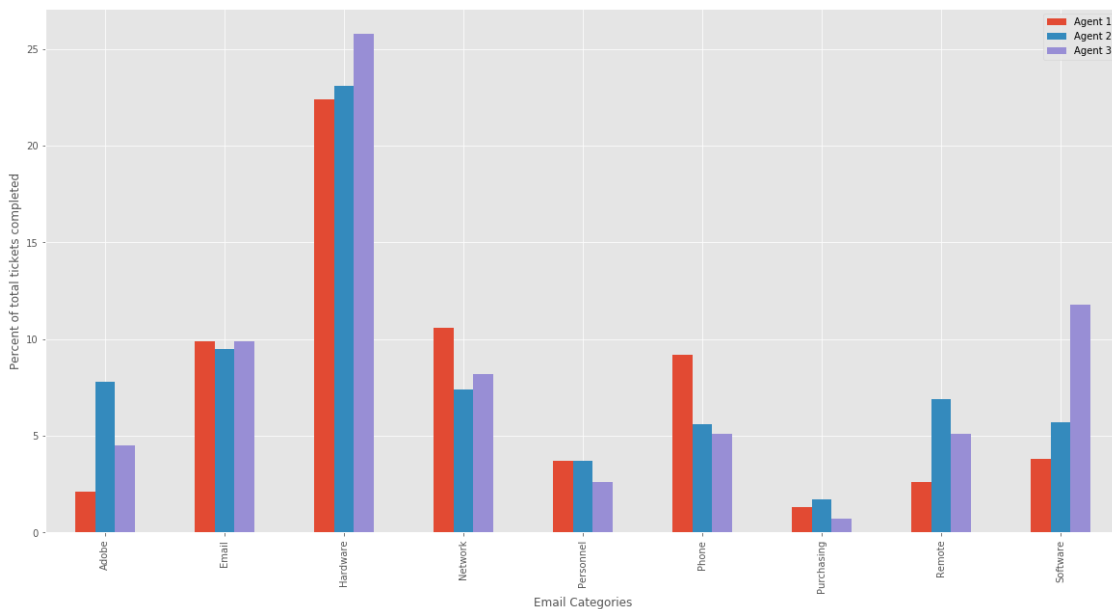
Training is one of the areas that help-desks might look to improve upon. We can use the style of data frame above to gather information about how effectively technicians have been trained,

and which areas of training may still need to be improved upon. Below are three Technicians. Agent 1 worked in the Helpdesk from February of 2016 to May of 2019. Agent 2 has worked in the Helpdesk from October of 2018 to present, and Agent 3 has worked in the Helpdesk from October 2019 to present.

[85]:

Email Categories	Percent - Agent 1	Percent - Agent 2
Adobe	2.1	7.8
Email	9.9	9.5
Hardware	22.4	23.1
Network	10.6	7.4
Personnel	3.7	3.7
Phone	9.2	5.6
Purchasing	1.3	1.7
Remote	2.6	6.9
Software	3.8	5.7
Percent of Total Inbox Solved	25.2	15.9

Email Categories	Percent - Agent 3	Percent - No Agent Specified
Adobe	4.5	3.1
Email	9.9	8.6
Hardware	25.8	18.7
Network	8.2	7.6
Personnel	2.6	1.0
Phone	5.1	8.1
Purchasing	0.7	0.7
Remote	5.1	2.3
Software	11.8	6.4
Percent of Total Inbox Solved	9.7	11.5



Due to these time differences, this can give us insights about training. For instance, we see that the percentage of Network tickets, tickets with keywords such as “internet”, “wifi”, “domain”, “permissions”, “drives”, and “network drives”, rises from Agent 3 and 2 to Agent 1. This shows us that with more time, agents may get more comfortable with network tickets or be asked to answer more of them. We see a similar trend with Phone tickets. On the opposite, we see a category such as hardware tickets, tickets including “station”, “setup”, and other computer and printer hardware words. We can see that this category gets smaller as the agents answer more tickets, this could show us that as agents are further trained, they are called upon less to answer tickets dealing with hardware.

Overall, the categories are fairly even, showing us that although help-desks may get more comfortable with certain types of tickets over time, **training is doing a good job of preparing new hires to deal with all types of tickets.**

To conclude our question about completed tickets broken down by category, we can say that we tend to complete tickets at about the same rate as we receive them in each category. Our biggest discrepancies are in the Hardware, Software and Remote categories. We can also make conclusions about our quality of training based on this breakdown. In particular, we can say, with a small sample size, that newer hires tend to complete tickets at about the same rate across categories that more experienced technicians do. We do see an increased amount of Hardware tickets and software tickets completed by newer hires than older hires, and more network tickets completed by older hires.

5 Discussion of Limitations and Future Considerations:

5.0.1 Limitations:

Like all analyses, the results can only be as accurate as the data allows them to be. These results are dependent on correct tagging of emails and the subject lines of our emails received. For instance, an email about hardware issue with the subject line “Help!” will not be counted as a hardware ticket in this analysis. The algorithm has been able to classify 60 – 70% of our tickets, which I feel is a large enough sample to be representative of the overall breakdown of our inbox, but this is still something to think about.

This analysis also relies on time that an email is sent. For instance, if someone has an issue with their email in the morning but doesn’t send the email until the afternoon, the email counts toward the afternoon ticket columns when it might be more useful to count it toward the morning.

Finally, one issue with breaking the analysis down by technician is that email tags which are removed from an Outlook mailbox do not stay on the tickets they were previously on. There were more technicians that I would have liked to be able to investigate to be more thorough, but their tags had been removed. Luckily, Technician 1’s tags were still in the mailbox, so I was able to analyze their tickets.

Overall, I feel that the quality of the data was sufficient for this analysis and that the results are not compromised by these issues, but they are something to be aware of nonetheless.

5.0.2 Future Considerations:

There are limitations to this Outlook data: things like matching up entire email threads are very difficult. In the future, there are some questions that we might still want to answer:

- Are our clients satisfied with the quality of our work?
- What is the help desk average response time?
 - By category
 - By time period
 - By technician
- What are the incidents resolved without escalation?
 - By category
 - By day
 - By technician
 - To identify training needs
 - To identify new tech requirements.
- What is our backlog duration?
 - How many tickets?
 - How long or what is the average time for a backlog ticket to then be resolved?
- What is our resolution time to unresponsive clients?
 - By ticket type
 - By technician