

STAA57 W21 - Project Proposal

Group 11 (Alice Huang, Dominic Ma, Jalal Kassab, Vanshika Virmani)

Link to the shared RStudio Cloud project that created this report: <https://rstudio.cloud/spaces/115177/project/2132391>
The final R code can be found in the Proposal_FINAL_Submitted.Rmd file.

Introduction

We are interested in efficient pricing of flight training courses and factors associated with student success. We plan to analyze costs that Durham Flight Training Centre must incur for offering flight training. We also plan to study combinations of exercises and aircrafts associated with success in training aviation students. We define students as ‘successful’ if they have flown solo at least once during the course of their training.

To analyze the costs of Durham Flight Centre, we would like to estimate:

1. How much money the training centre has spent on fuel for different aircraft models over the past few years.
 - Durham Flight Centre’s website also lists the hourly rental rate for each plane. This information will help us estimate revenue and determine the profit margins for the flight centre.
2. Whether the number of flight hours have differed across different times of the year. And if some planes are more in demand during certain months.
 - This will help us determine how demand for flights varies throughout the year. If we find that certain planes are more in demand during certain months, then we can analyse how the flight centre can adjust the prices for training with different aircrafts to increase profit.
3. How much money it costs to train the average student until they can safely take off, land and fly solo.

To answer questions related to training of students, we would like to build a “profile” for students. In order to determine how the centre can maximize student success, we want to know:

4. What are the types of exercises that students who were able to fly solo at least once completed that set them apart from students who weren’t able to fly solo at all?
5. Which types of planes did the ‘successful’ students use to complete certain exercises? This will help us map out information for the Durham Training Centre, as to which planes they should allocate for which exercises to maximise student success in training.

By further studying the types of exercises and aircraft models suited for those exercises, we can determine how the training centre should schedule certain exercises for students. More student success can help the centre attract more students.

Data Analysis Plan

1. **Monthly Training Duration:** Grouping the sessions by month and summing up the duration would show us the total time flown per month. This would allow us to see if demand for flight training is influenced by season.
2. **Profit:** Summing up the training duration and multiplying that by the costs of fuel and instructor salaries would give us the total costs. Multiplying the sum of training duration by the hourly rental rate of each plane would give us the revenue. The profit would then be the revenue minus the costs. We can plot the type of planes vs the total cost for each plane, and obtain insight into which plane(s) are the most profitable.
3. **Costs of Training Average Student:** To estimate the cost for the “average student” in the training centre, we will group the data by different student IDs and plane types and then find the average number of hours spent in flight training by each student for each plane type. We will then multiply these values with the average costs from above, and create a graph visualizing our findings.
4. **Successful students:** First, we will see which students flew solo at least once, by filtering for rows where training type contains the substring “solo”. Then we will graph the frequency of exercises completed by each student to see the most frequent exercises completed by “successful” students. We’ll repeat the same process with filtering out students who got to

fly “solo”. We will then add up the duration of hours each successful student flew with different types of planes to see if successful students used certain planes more frequently than others. Grouping the most frequent exercises completed by type of airplane will show us if certain aircrafts are better suited for particular exercises.

Data

Our primary source of data is a dataset provided by Durham Flight Training Centre. This dataset logs the exercises students completed with different instructors on different dates. It specifies durations of the training sessions, whether exercises were completed solo or with an instructor, and which aircrafts were used. The dataset only spans the years 2002, 2015-2020, so we can only observe patterns and compare trends over the span of five years. It may not be suitable for generalizing trends to longer periods of time. We decided there was more reliable data for 2018-2020 than the other years so we focused our analysis on these years. We used the Durham Flight Centre’s website: <https://durhamflightcentre.com/> to get hourly aircraft rates, flight training package rates, and instructor rates.

We gathered data on the aircrafts’ ownership costs from aopa.org, the website of the “Aircraft Owners and Pilots Association”, and <https://cessna150152club.org/>, the website of the “Cessna 150-152 Club”, a membership club and nonprofit dedicated to educating prospective and current pilots about the Cessna-152 and Cessna-150. We believe these organizations’ numbers should be reliable since these organizations and communities have access to active users of Cessna-172, Cessna-152, Cessna-150 planes. Whenever sites provided a range of prices, we just used the median for our computations as the median is a good indicator of the “middle value” in a distribution and is less likely to be affected by outliers. We used the Tillsonburg airport website (<https://www.tillsonburg.ca/en/live-and-play/Fuel.aspx#>) to find the cost of Avgas (the same type of gas the training centre’s aircraft use). In April 2021, it said Avgas was \$1.69/L. We believed this was an accurate estimate for fuel pricing, since Durham Flight Centre is in the same province as Tillsonburg airport so region and market conditions affecting fuel prices should be similar. See Appendix for more information about our data sources.

Importing and Formatting the Data in R

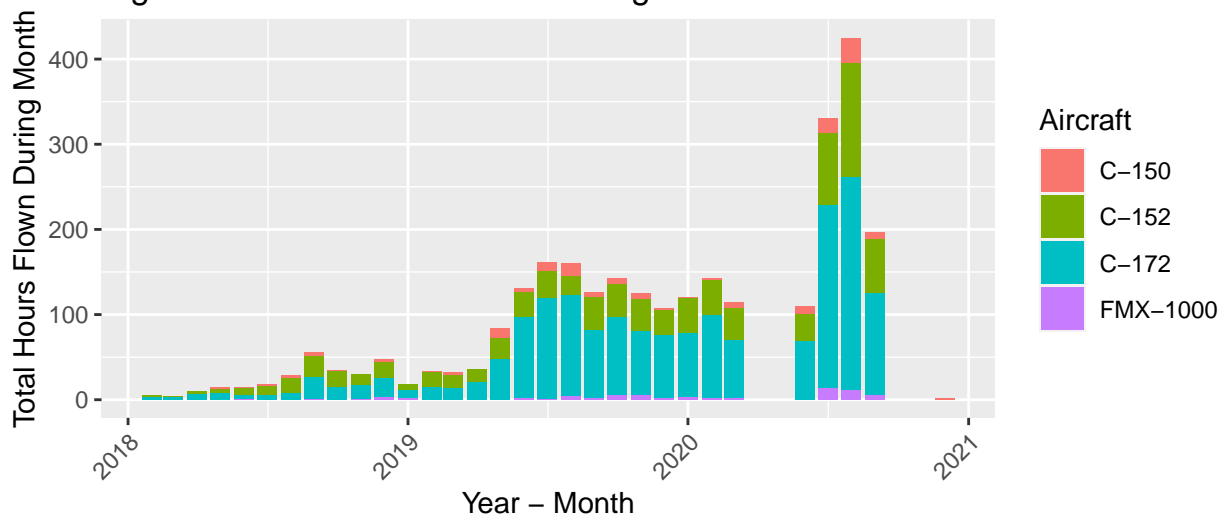
Summary of Changes Made to Dataset We made changes to the dataset, with the goal of amending input errors in the data, and making our analyses and visualizations more accurate. For example, we removed entries where the Session_ID’s were repeated, which would have gone against the intention of having them be unique identifiers. We also decided to remove rows in which Aircraft was blank, since much of our analyses relied on this variable. Finally, we modified any entries that were obvious typos, like “C152” and “111” which were meant to be “C-152” and “11” respectively. See Appendix for more details.

Analysis

Training Duration

We wish to estimate the operating and maintenance cost the training centre has spent on each type of aircraft over time. Thus we are interested in the total amount of time students spent training with different types of aircrafts. Computing the hours spent on flying the different aircrafts will help us estimate how much fuel was consumed by each type of aircraft, and in turn, how much money was spent on fuel for each aircraft. We graphed the months of the year against the total duration of training session for those months.

Figure 1: Months vs Duration of Flight Hours



Note that data for certain months is missing. 2020 only has data for 7 months because of the lockdown from the pandemic. 2018 has no data for January. Finding the aggregate duration for each month across all years (eg. adding up duration for 2018 January, 2019 January, etc) would yield skewed results. So we decided to treat months for different years separately.

Training duration generally increases during the months of June, July, August, September, and October, usually peaking around August. This is most likely due to the fact that those months tend to be warmer, and invite clearer skies which are an important consideration when flying. We think it'd be interesting to compare the training duration with more detailed weather data on temperatures and precipitation. Moreover, we see that the hours peak around 60 in 2018, around 180 in 2019, and oddly enough, 2020 seems to have the highest peak of over 400 flight hours. So even during a pandemic, the flight center was able to maintain growth and bring in students to log more flight hours. This steady growth in flight hours, which is synonymous to the demand for training, has been increasing over the years, as shown in our graphs. But to help tackle the slump in demand that the flight center sees every winter or spring, the centre could consider decreasing prices for the training courses or give special package rates (of course while staying above the fixed + variable costs), which in turn could possibly increase demand and growth in the future.

Plane Fuel Consumption

We wanted to compute the flight centre's profits so we wanted to estimate various costs, one of those being fuel costs. We found estimates for how much fuel each plane burns, in gallons per hour - the unit pilots conventionally use to measure hourly costs. For each training session, we computed the cost of fuel in gallons per hour and multiplied it by the cost of 1 gallon of Avgas (\$7.68/gallon). See the Appendix for more details about our fuel cost estimation process.

Aircraft	total_fuel_consumed_gallons	price_total_fuel_CAD
C-150	936.88	7197.95
C-152	5238.07	40243.54
C-172	14890.68	114403.51

We found that over the years 2018-2020, the training centre's Cessna-172 aircrafts consumed around \$114403.51 worth of fuel (14890.68 gallons) in total. The Cessna-152 aircrafts consumed around \$40243.54 worth of fuel (5238.07 gallons) in total. The Cessna-150 aircrafts consumed around \$7197.95 worth of fuel (936.88 gallons) in total. This is consistent with the fact that during 2018-2020, the Cessna-172 was used for training more often, and its hourly fuel cost is more expensive than that of the Cessna-152 and Cessna-150. It is interesting to note that the training centre spent nearly 2 times the amount of time on Cessna-172 than the Cessna-152, but 3 times the amount of fuel. This suggests that the C-172 is a more expensive plane to operate. The flight centre could consider offering more special packages with the C-152 & C-150 and decreasing the hourly rates for students, with the aim of increasing demand for C-152 & C-150.

Monthly Instructor Costs

Another cost incurred by the training centre is that of paying instructors for dual training and ground school. We found the duration of training sessions with "LF-dual" training type and "GROUND" Aircraft type. For sessions involving "GROUND" Aircraft type, we noticed that the duration was entered as NA (not given) so we didn't know how many hours the instructors were being paid to teach Ground School sessions. We approximated the duration of each ground school session by comparing it to the average durations of other sessions in which students completed the exact same combination of exercises. We multiplied the approximate hours of ground training by the hourly instructor pay rates listed on Durham Flight Centre's website (\$65 (CAD)). See Appendix for more details.

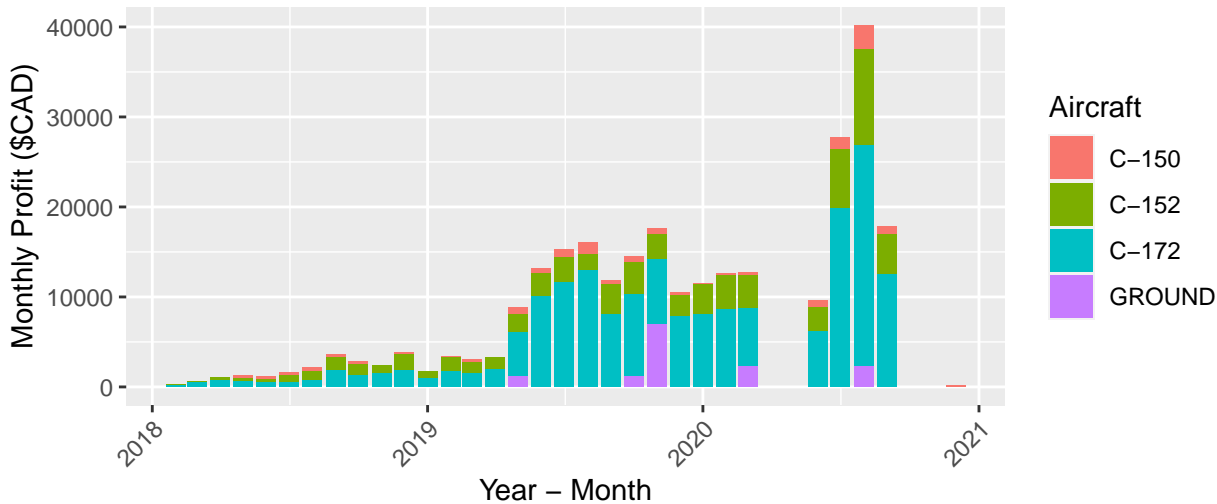
Monthly Revenue

We estimated the flight centre's revenue using the durations of training, and the hourly costs of renting planes listed on the flight centre's website. We didn't know how much the flight centre earned from Ground School alone, so we estimated it using the flight centre's package rates. See Appendix for more details on our assumptions.

Monthly Profit

The two main costs we examined were fuel costs and instruction costs, since computing them was more accessible given the flight centre's dataset and website pricing information. We computed the profit by subtracting these costs from the revenue, and graphed it by month. Operating a flight centre and offering flight lessons would certainly incur other costs, however we didn't have enough reliable information for these costs. If in the future we acquire such information, we would certainly like to update our profit estimates.

Figure 2: Year – Month vs Monthly Profit

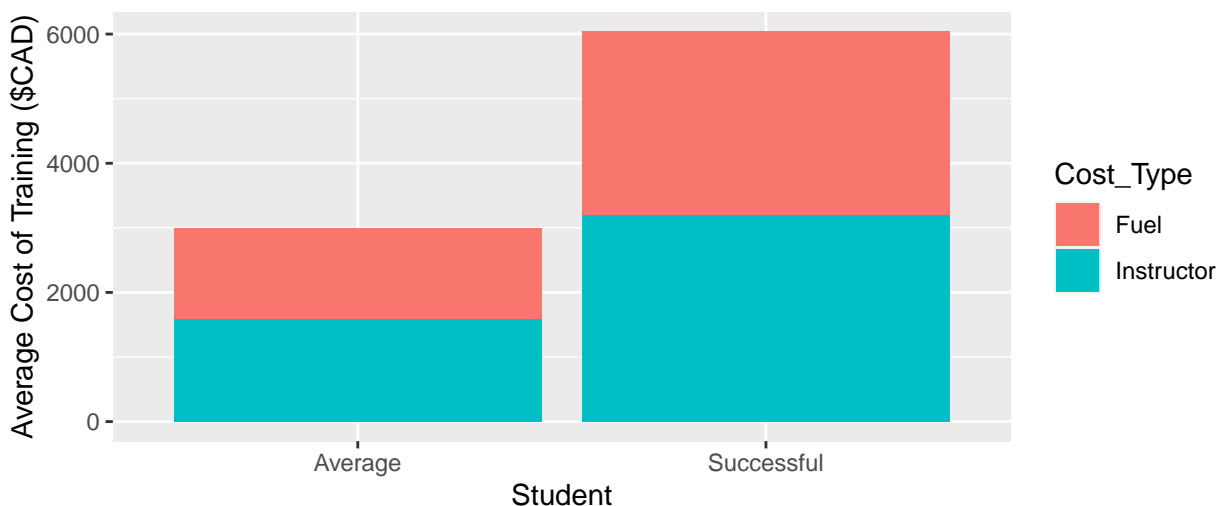


It seems like most of the flight centre’s profit came from renting the C-172 planes to students during the months of July, August, and September, even though Instruction Costs and Fuel Costs for C-172 were also higher. This is likely because students spent more time flying C-172 planes and the C-172 planes were more expensive to rent.

Cost of Training Average Student

We wanted to estimate the fuel cost and instruction cost of the average student at the flight centre. We were also curious, out of the students who successfully flew solo, what was the average amount of money the flight centre spent on their fuel and instruction?

Figure 3: Costs of Training Types of Students



We estimated that the flight centre spends \$1406.75 on fuel and \$1590.28 on instruction for the average student in their program. The flight centre spends \$2843.00 on fuel and \$3197.82 on instruction for the average successful student in their program. The average successful student requires nearly 2 times the fuel and instructor costs than the average student. This makes sense, since the average successful student likely spends more time training than the average student. The average successful student will likely require more fuel for more aircraft practice, and more Instruction time.

Factors Common To Successful Students

We made two bar graphs. The first shows the students who were able to fly solo at least once, their aircraft choice, and the total time flown on those aircrafts. The second shows the same information for students who were NOT able to fly solo at least once. These graphs show us the “successful” student’s behaviour and their aircraft of choice.

We saw that those who have flown solo prefer sticking to one aircraft and rarely switch aircrafts. We believe that students who stick to their preferred aircraft end up becoming more comfortable with it and mastering it at a faster pace. Meanwhile students

who tend to switch between 2 or 3 aircraft might struggle with adjusting to different aircraft, and end up taking a longer time to perfect their craft.

Figure 4a: Successful Students vs. Flight Time on C-152, C-172

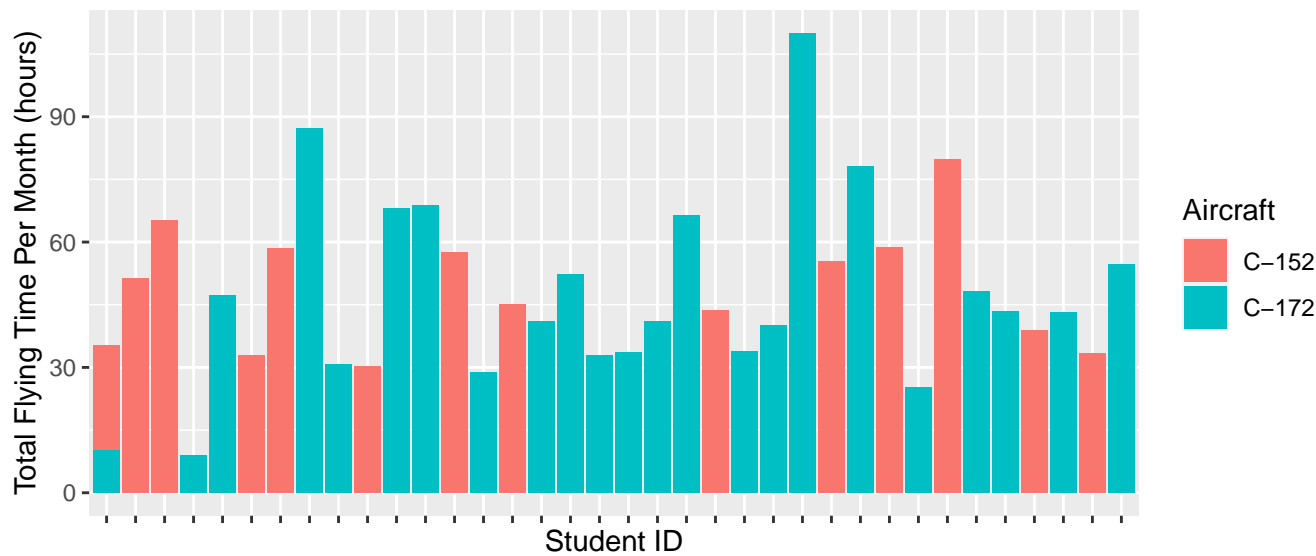
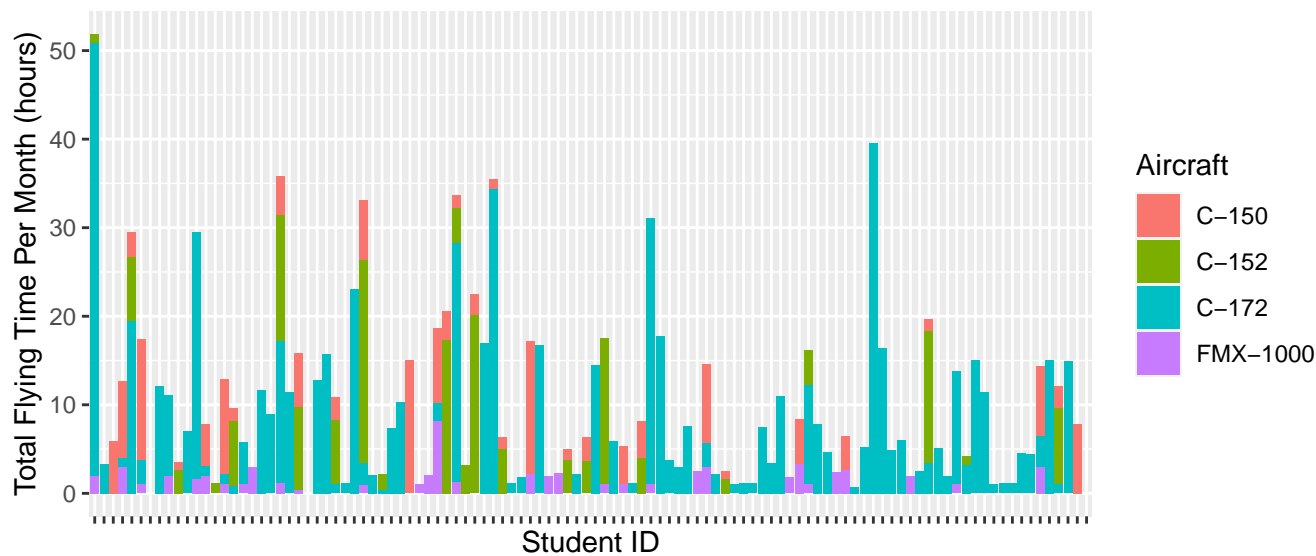


Figure 4b: Unsuccessful Students vs. Flight Time on Aircrafts

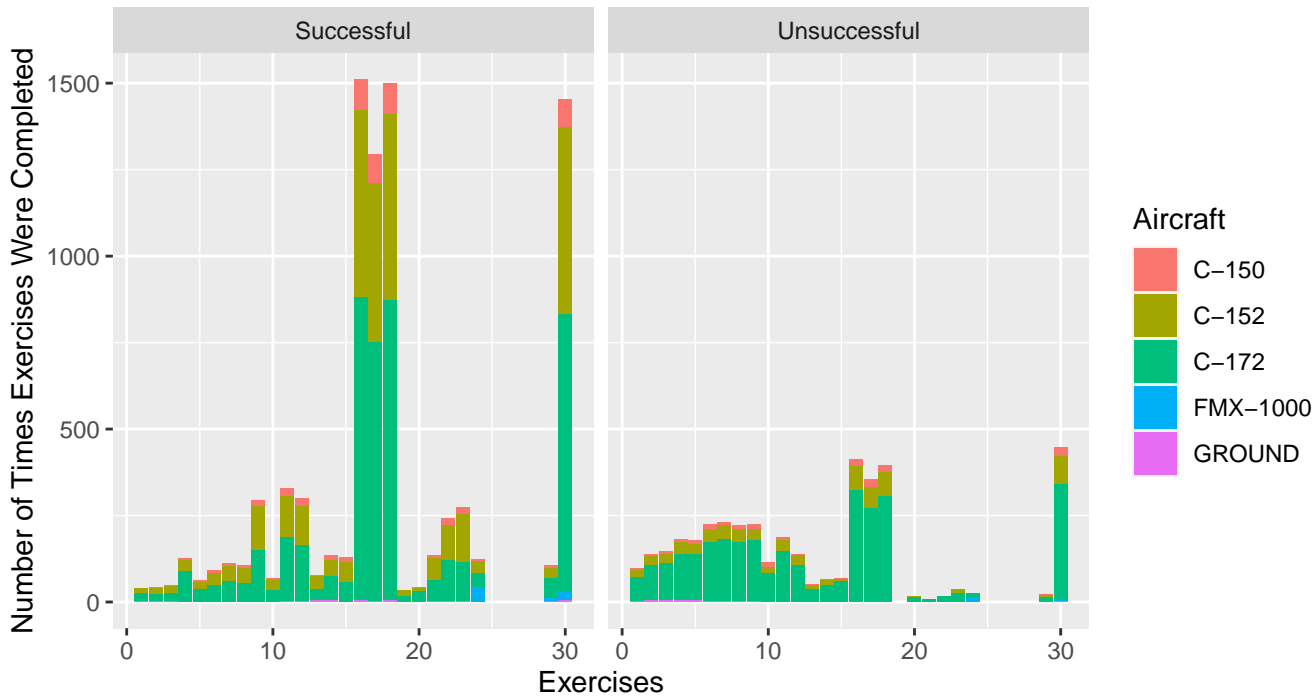


Successful Students Exercises

We were interested in what successful students do that sets them apart from unsuccessful students. On the next page, to the left, we have a plot showing the frequency of Exercises completed by all successful students combined. On the right we have a plot showing the frequency of Exercises completed by all unsuccessful students combined.

We see that the successful students and unsuccessful students completed Exercises 16, 17, 18 and 30 the most often. Exercises 16, 17, 18, and 30 respectively correspond to Take-off, Circuit, Approach and Landing, and Radio Communications. All the successful students combined completed Exercises 16, 17, 18, and 30 1512 times, 1293 times, 1499 times, and 1453 times respectively. All the unsuccessful students combined completed Exercises 16, 17, 18, and 30 412 times, 354 times, 396 times, and 448 times respectively. Successful students practiced these exercises over three times as much as the unsuccessful students. This suggests that Take-off, Circuit, Approach and Landing, and Radio Communications are key components of pilot training. We believe that the number of times a student practices these exercises may be correlated with ability to fly solo. We also noticed that successful students completed Exercise 29 (Emergency Procedures) nearly 5 times as often as unsuccessful students. Successful students also completed Exercises 21-24 (Precautionary Landing, Forced Landing, Pilot Navigation, Instrument Flying) more often than unsuccessful students. Successful students likely have more practice doing different types of landing and navigation than unsuccessful students.

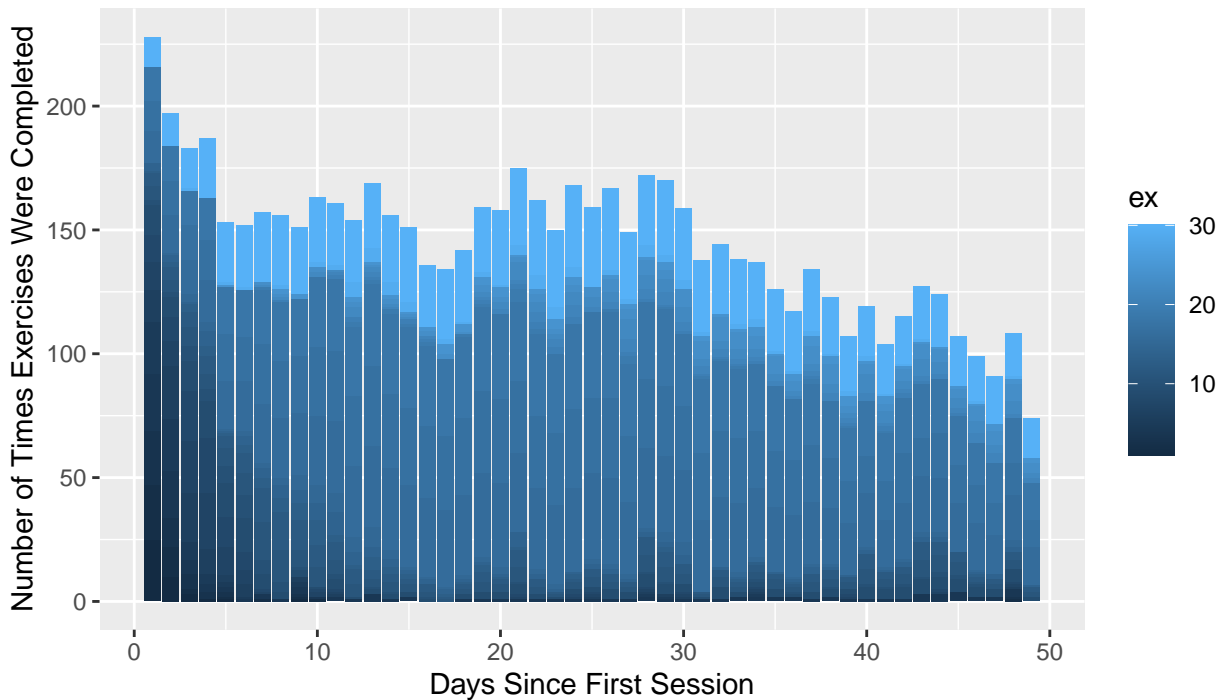
Figure 4c: Number of Times Exercises Were Completed



Progression of Training

We were also interested in finding out what the progression of a successful student's training looks like. Here is a graph showing the frequency of exercises successful students complete as their training progresses. We found that the average successful student's training lasts 50 days, so we showed the frequency of exercises for the first 50 days of training.

Figure 4d: Number of Exercises Completed Over Training



Here we see the progression of training, as successful students mostly focus on familiarisation and preparation for flights (through exercises 1-5) for the first couple of days, and shift to slightly more advanced training with maneuvering and controls (through exercises 6-10) in the next few days. From Day 11 onwards, we observe that exercises 16, 17, 18, 30 are practiced often and continue to be practiced throughout the training. This is indicated by the medium blue color that dominates a large area of the graph. Students may sometimes practice Exercises 20-24, and 29, involving Illusions Created by Drift, Precautionary Landing, Forced Landing, Pilot Navigation, Instrument Flying, and Emergency Procedures, which are more advanced pilot training exercises.

From graph 4c, we see that even though all students practise the more basic exercises (1-10), unsuccessful students practise these in greater frequency. The more prominent difference between the two categories of students, however, stems from the observation that successful students practiced exercises 16, 17, 28, 30 over three times as much as the unsuccessful students. We also observe that the proportion of successful students that use C-152 for completing these exercises is much greater than the proportion of unsuccessful students, who tend to mostly stick with C-172.

One can infer that successful students display high dedication and consistency in completing Take-off, Circuit, Approach and Landing, and Radio Communications throughout their training. But it would be misleading to conclude that students who shift their focus towards exercises 16, 17, 28, 30 or the ones who train with C-152 are more likely to become successful, because we cannot assume that all students have the same level of experience before they start training at the flight centre. It is possible that some successful students have prior experience in flying, and that is why they spend less time on the more basic exercises. On the other hand, someone with less or no prior experience should focus a great deal on the basic exercises to familiarise themselves with the more basic exercises and getting comfortable with one type of plane.

We hope that these student profiles of ‘success’ will assist the flight centre in designing training schedules that offer the best value-for-money sessions to the students. Depending on the student’s experience, the flight centre can recommend which exercises are best suited for the student and how many times they should practice before progressing to more advanced ones. They can also recommend which plane is more suited for completing each exercise, while considering the student’s familiarity and preference.

From the data and graphs, we didn’t observe a ‘hard-and-fast’ rule to achieve success. Even though we don’t have data about the student’s prior experience in training, we can only infer that students who wish to be more successful should simply dedicate more hours in training. The flight centre can use this data to encourage more students to apply more effort and persistence through longer periods of training to become successful.

Summary

Overall, we found that the flight centre’s students log more flight hours during late summer to early fall months, and prefer to do so on the C-172. The centre earns more revenue and profit during these months on C-172 aircrafts, but also spends more on fuel and instruction. We computed fuel and instruction costs of training the average student and compared them with the same costs for the average “successful” student. We explored common traits of successful students and determined that successful ones seemed to stick to one aircraft consistently. We graphed the progression of successful students’ training. Exercises 16, 17, 18, and 30 were important in student success. We hope our findings are interesting and relevant to Durham Flight Centre.

Appendix

References

We consulted the following sources to learn how to process certain types of data and format R Markdown files.

<https://e2a207d9ab5c4c3d9007aa1ad8e21ccd.app.rstudio.cloud/help/doc/html/packages.html>

<http://www.stat.cmu.edu/~cshalizi/rmarkdown/>

<https://stackoverflow.com/questions/29568956/how-to-add-a-rank-column-to-sorted-dates-in-r>

<https://bookdown.org/yihui/rmarkdown/pdf-document.html#data-frame-printing-1>

https://rmarkdown.rstudio.com/authoring_basics.html

More Information About Data Sources

The data gathered from the “Cessna 150-152 Club” may be a little biased since the entire organization is based on those two planes and the content on the website seems to feature more positive anecdotal experiences with the two planes. Furthermore, the ownership club may not include as much information for users who did not pay their club membership fee, and others who are not active on their internet forums. Due to these concerns, we checked if their numbers for ownership costs were consistent with other websites, and they were, so we chose this as a data source.

The website for the “Aircraft Owners and Pilots Association” helped us find out the fuel efficiency and price for all three planes. But we had to decide whether we should use the average used price of the planes or the price of the “reimagined” planes, which have been overhauled, repainted and serviced. We ended up using the price of the “reimagined” planes since it gave us a better idea about the actual price difference without any external variables.

Both the “Aircraft Owners and Pilots Association” and “Cessna 150-152 Club” are based in the United States so costs, regulations and experiences may be biased towards English-speaking American users of Cessna-172, Cessna-152 and Cessna-150 planes. Our

client is based in Southern Canada, close to the US border, but we believe certain costs like a plane's fuel burned per hour, should not depend on region. Insurance and inspection costs may vary by region, however.

Data Processing Methods and Details

Data processing explanation: The goal of this script was to make changes to the main data object that we will all be working with. We call the object `my_cd`. The first step is to remove rows with duplicated "Session_ID", since these are meant to be unique. This was done using the `distinct` method. Next, we changed the exercise column from being string into a list of character vectors using `str_split`. This was done so that we could perform counting and sorting operations, as well as catching typos, easier.

Next, we used standard subsetting to change typos "C152" in Aircraft, and "111" in Month, into "C-152" and "11". We were able to determine that "111" was meant to be "11", since the original data on the spreadsheet was ordered; what appears to have happened was a "1" was taken from the previous entry and added to this one. We ensured all Cessna-152 planes were entered as "C-152" for consistency. Also, the choice to use standard subsetting was appropriate since the change would not be misapplied to other data; the search condition only applies to what we want. In the next step however, we needed to use more specific criteria, so we used the "replace" function within "mutate" to fix the affected Date columns, and the "1" in month that was meant to be "11".

Next, we removed all entries in which Aircraft wasn't given, since much of our analyses rely on this information. This was done by filtering the complement of rows for which `is.na` evaluated true. Lastly, the bulk of errors and typos seemed to have come from the exercises column. This step was particularly tricky, since our instinctive approaches did not work. Beginning with empty string which appeared as `c("")`, and mostly arose from double commas in the original data. Initially, we tried doing: `my_cd %>% filter(str_detect(Exercises, "~$"))` to find instances of empty string, and then replacing them using something like `x[!(x %in% "")]`. In the end, the method that worked was combining the `lapply` function, which applies a function to each element in a list, along with the `stri_remove_empty` function that removes empty string from a character vector.

After that we used conditional subsetting again to replace individual occurrences of errors. `str_detect` generates a logical vector for each error, and `which` returns the indices of rows that evaluated TRUE. We then used those indices to manually replace each character vector in the list. Finally, the row with "40" was removed, simply because it didn't resemble anything valid and it was the only one.

Airplane Costs Data We found estimates for fuel and operating costs from the "Aircraft Owners and Pilots Association" and the "Cessna 150-152 Club". These seemed to be the most accurate numbers we could find. These websites did not post their numbers as datasets with csv, txt, xlsx, xml, or other formats that are convenient to work with. In particular, on the "Cessna 150-152 Club" website's "Members Only" pages, it says that servers containing their data were hacked, so the organization had to pull their data offline and is now in the process of "building a modern and secure database structure". Since we could not find a pre-made dataset, we made a csv file, and entered the hourly estimates of various costs for operating and maintaining the aircraft in the tidy data format. We put this csv file in our data folder with the rest of our project.

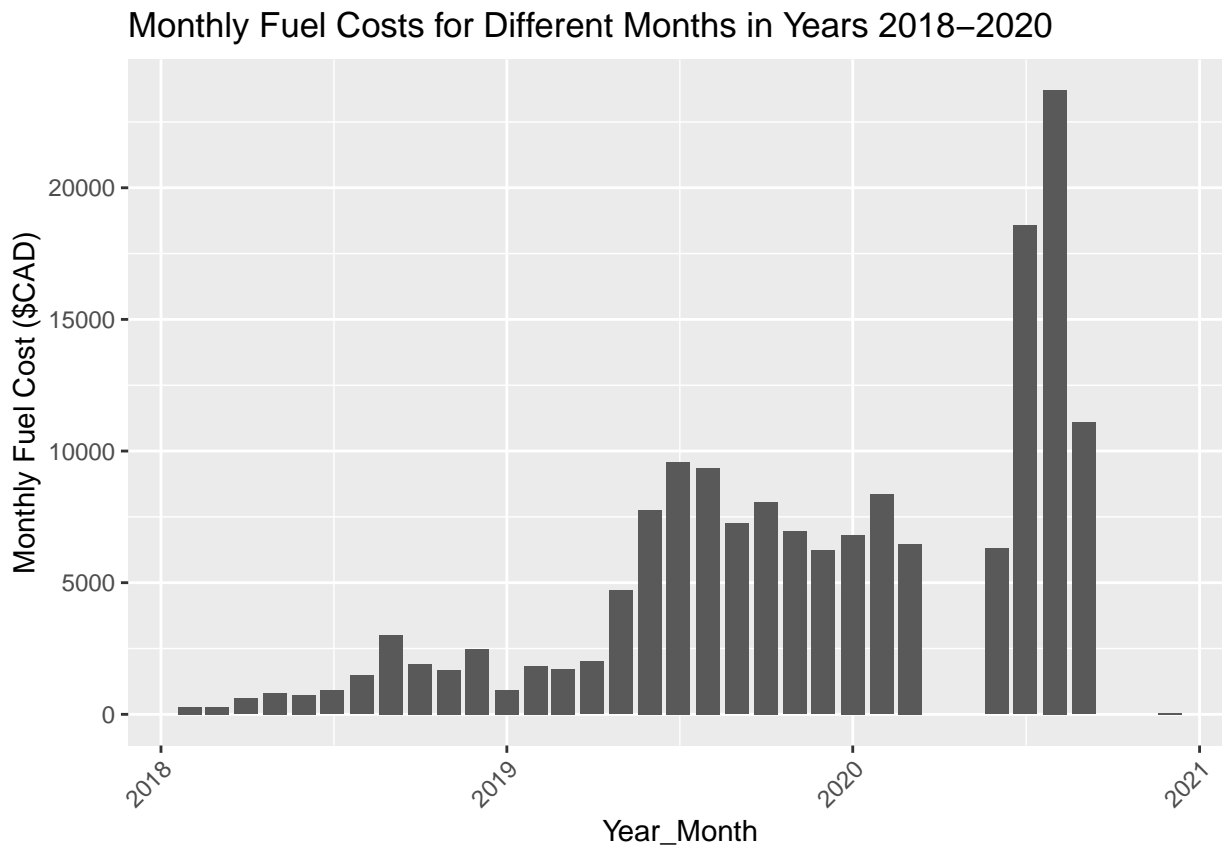
We found the cost of Avgas (the same type of gas the training centre's aircraft use) in \$CAD/L from the Town of Tillsonburg's website (<https://www.tillsonburg.ca/en/live-and-play/Fuel.aspx#>) under the sections Airport > Services > Fuel. It said Avgas Fuel Prices were \$1.69 per litre. We thought this was the most accurate estimate we could get for fuel pricing in dollars, since Durham Flight Centre is in the same province as this airport so region and market conditions affecting fuel prices should be similar.

We used the conversion factor 1 Litre = 0.2199692 gallons. So \$1.69/litre is approximately \$7.68/gallon. Multiplying gallons of Avgas consumed for each session by the price of gas per gallon gave us the price of gas consumed for that session. We made a column that multiplies the number of gallons consumed by its corresponding price depending on plane used. We put a corresponding graph in the appendix for the sake of space.

A limitation of our current analysis is that we assumed the cost of fuel consumed per hour remained constant over 2018-2020 regardless of aircraft model, age and condition, nature of exercise, market conditions for fuel pricing (especially given COVID-19 pandemic), and other factors. We assumed FMX-1000 and Ground training consumed 0 fuel.

When we looked at the fuel consumption per year, we found that in 2019-2020, the Cessna-172 consumed much more fuel than the Cessna-152. In 2019, the C-172 consumed approximately 6622.56 gallons while the C-152 consumed around 2009.95 gallons of fuel. In January-March, June-September 2020, the C-172 consumed approximately 8106.84 gallons while the C-152 consumed approximately 2940.20 gallons of fuel. Again, this is consistent with the frequency of Cessna-172 training sessions during those years.

Monthly Fuel Costs Graph



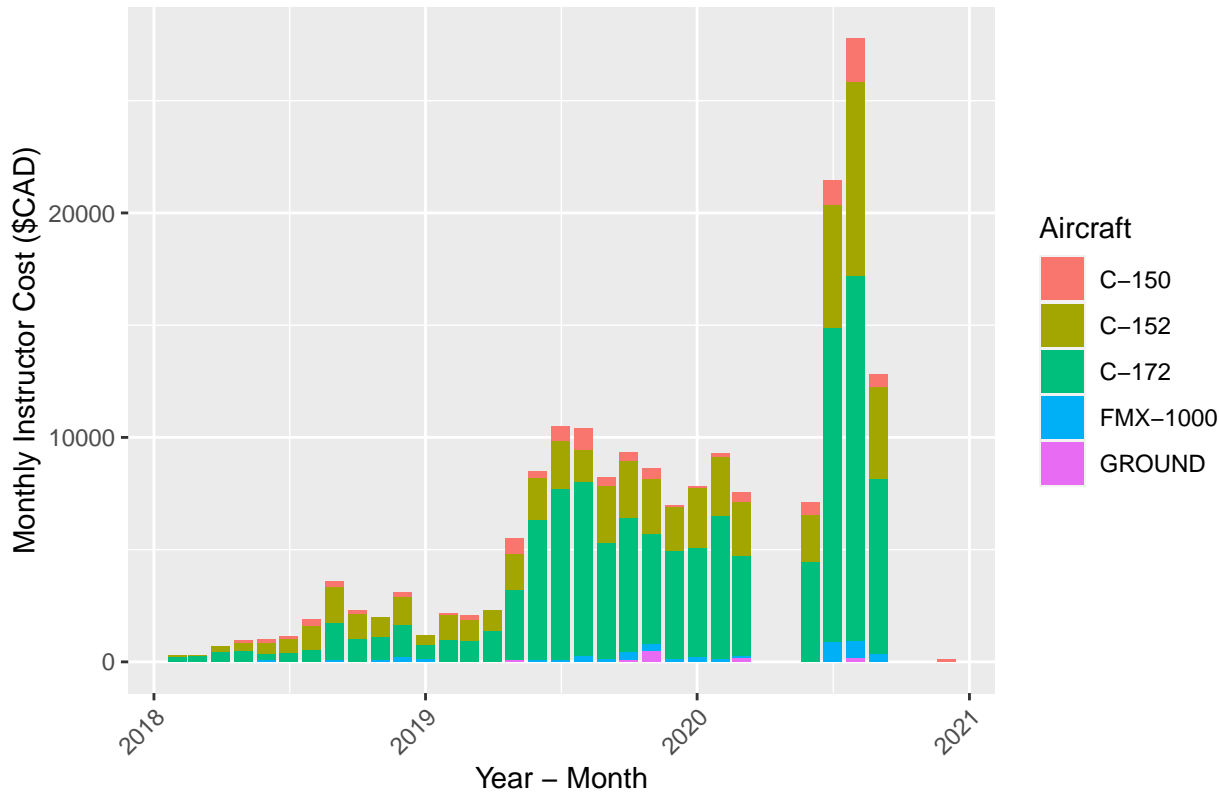
Monthly Instructor Costs: More Information

Both of these correspond to when the instructor was involved in either pre-flight or flight training with the students.

For the “GROUND” Aircraft type, we noticed that the duration was entered as NA (not given) so we didn’t know how many hours the instructors were being paid to teach Ground School sessions. Initially we thought of completely ignoring the Grounded Aircrafts and the instructor fee for these sessions, but ground training makes up a significant cost of the packages that students pay for. And completely disregarding this fee would highly skew the data. So we approximated the duration of each session of ground trainings by comparing it to the durations of other sessions for which the students completed the exact same combination of exercises. We then averaged those durations to approximate the hours of ground training each student trained for. From here, we created a column corresponding to the approximated ground training hours for Grounded Aircraft type. And we combined the two tables (original and grounded) to have this approximated duration for grounded planes (on the rows that originally showed N/A). On the Durham flight centre website, the hourly rates for instructor fee are listed as \$65 (CAD). So we multiplied the duration of these ground trainings and LF duals with the instructor fee and grouped the data by Month_Year and Aircrafts to show the monthly cost of instructors for each type of Aircraft.

The graph of instructor cost is in line with the graph of fuel cost. This makes sense because the cost is synonymous with the duration the planes are flown for.

Year_Month vs Monthly Instructor Cost



Monthly Revenue Assumptions

We wanted to estimate the flight centre's revenue with the ultimate goal of estimating profits. We didn't know how much the flight centre earned from Ground School alone. On top of this, due to COVID-19 pandemic, the flight centre also stopped offering in-person ground school sessions. Due to this ambiguity, we estimated revenue using the training duration from the flight centre's provided dataset and the flight centre's website showing the different pricing options. We assumed these prices were constant throughout 2018-2020.

We assumed each GROUND student purchased one package because we thought each student would only do preparatory GROUND training when they are first learning to fly, and not throughout their pilot education career. When we looked at the data, we saw that students who did GROUND training had only one or two rows corresponding to GROUND training with dates very close to each other.

We assumed the revenue earned from GROUND training was equal to the amount of revenue earned from a package minus the amount of revenue earned from flying in the aircraft. We assumed that the revenue earned from students flying in aircraft on hourly rental and students flying in aircraft on package deals was the same. We compared the revenue earned from GROUND training for the 5-hour and 10-hour package, and took the higher of the two to get a higher upper bound. We also assumed a person who does a 10 hour package probably won't do more ground school than a person who does a 5-hour package if the end goal is to fly aircraft solo. For students who chose the 10-hour package, we assumed the ground school cost the same for students who flew with C-152 or C-172. Then for the rest of the training of students who took the packages, we decided to check what plane they used and assumed the revenue earned was proportional to the flight rental hourly rates.

During our revenue calculations, we rounded up the duration of hours flown to a whole number. For example, if a student flew 2.3 hours, we assumed that the flight centre charged them for 3 hours.

Year_Month vs Monthly Revenue

