STAA57 W21 Draft Report

Group 14 (Alec Larsen, Aditya Kulkarni, Vishal Sahoo, MD Wasim)

Link to RStudio Cloud shared project: https://rstudio.cloud/spaces/115177/project/2422226

Introduction

The profitability of a flight school business is reliant upon two critical factors: The efficiency of flight training and the success of students post that training. The two aspects of training we will focus on are: - Scheduling: We aimed to determine whether training sessions are being planned efficiently, and if not, how can they can be. - Student Success: We determine the ideal conditions that lead to student success. We define student success as completion of the first solo flight and subsequently, the completion of the course. We conducted our analysis through a series of graphs, tables, and statistical tests to prove or disprove our predictions.

Data

In answering these questions, we made the assumption that all students begin their training at the same skill level. Additionally, we assume there were no unmeasurable external factors impacting the duration or efficiency of sessions. That is to say, unless a factor can be measured, it cannot be accounted for in our analysis.

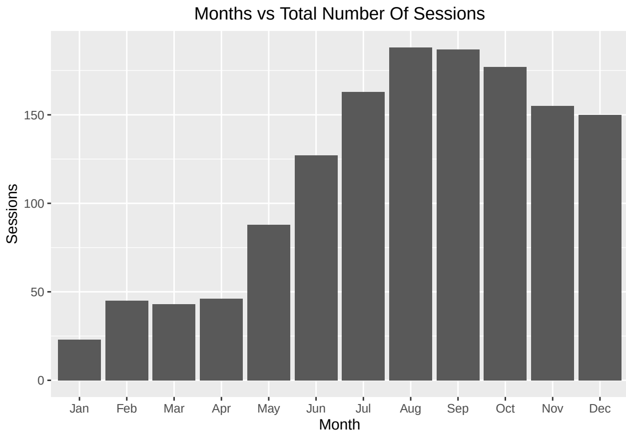
To determine when flying is the safest (least wind speed), we used weather data from Government Of Canada's Website.
Weather data: https://climate.weather.gc.ca/historical_data/search_historic_data_e.html

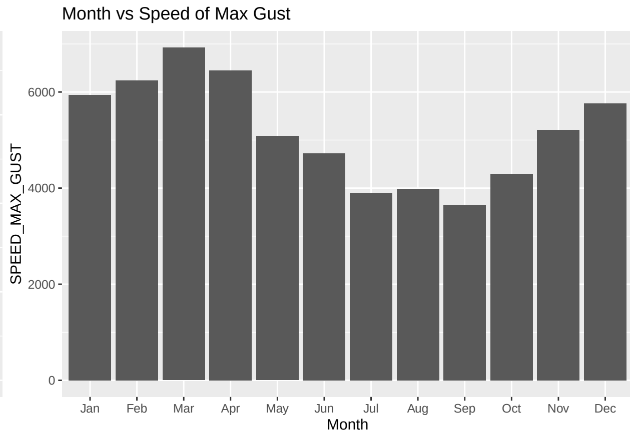
This data was used to compare the weather conditions during training sessions that are currently being held to possible ideal conditions. We will utilise the information stored in the following variables: Year, Month, Day, SPEED_MAX_GUST

Air traffic data was also used to determine the density of air traffic per month. To help determine when air traffic is the lowest, and thus flight accident risk is the lowest, we acquired air traffic data from the suggested supplementary data.
Air Traffic - https://open.canada.ca/data/en/dataset/b91772ed-edae-4fd4-8b80-a3e4c1d29976
We utilise the following variables from air traffic data: Month: Month of year, Value: The number of airplanes.

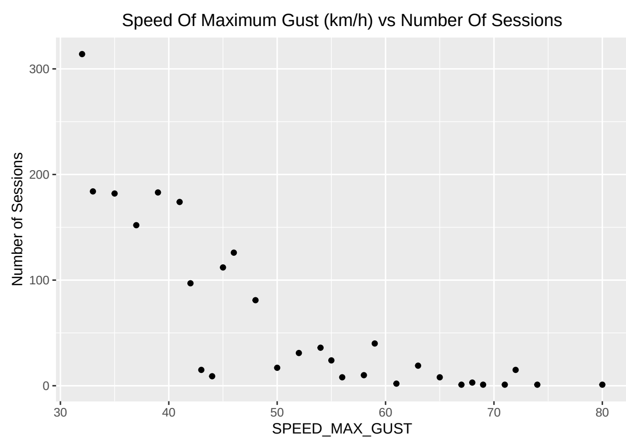
Analysis

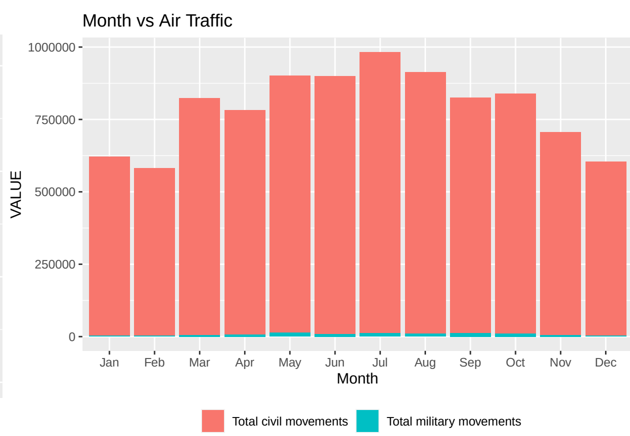
Scheduling





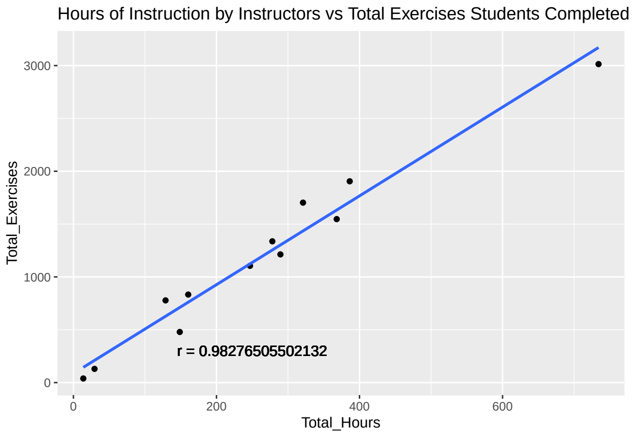
The data comparing the maximum speed of gust to the month of a year from 2011 to 2019 shows a trend line steadily increasing from January to March and then decreasing from March to the least in September. An increase is then observed till December. The number of sessions conducted per month for the given date range suggest that the conduction of sessions is least in the month of Januaru from where it starts rising and reaches its maximum in August. There is a steady increase from August up to December. A dramatic increase is observed in the month of May.

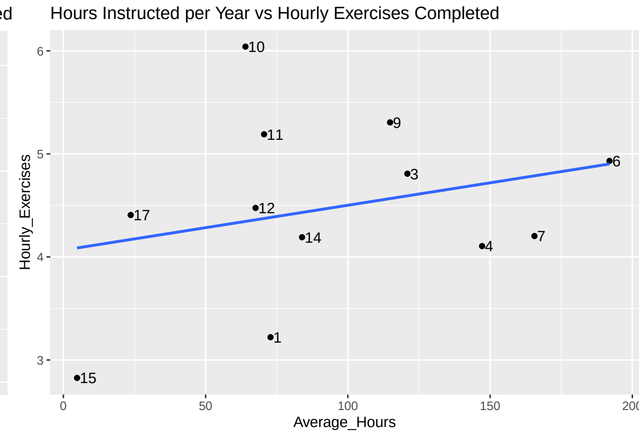




On comparing the speed of max gust to the number of sessions being conducted during that day, we found out that as the speed of gust increases, the Number of sessions being conducted are reduced. The Air traffic per month also has a trend with an overall rise from January to the maximum in July and an overall decrease till December. The months from May to October seem the best time to fly during the year. As the Air traffic is highest in July, and the number of sessions conducted are highest in August, the flight training is safety compliant already. Considering all factors, September could be the most ideal month to increase flight training.

We continued our analysis by establishing a reliable measure of instructor efficiency. As both the number of exercises completed and the number of hours each instructor works are continuous variables, we chose to establish an efficiency metric based on the correlation between these two variables.

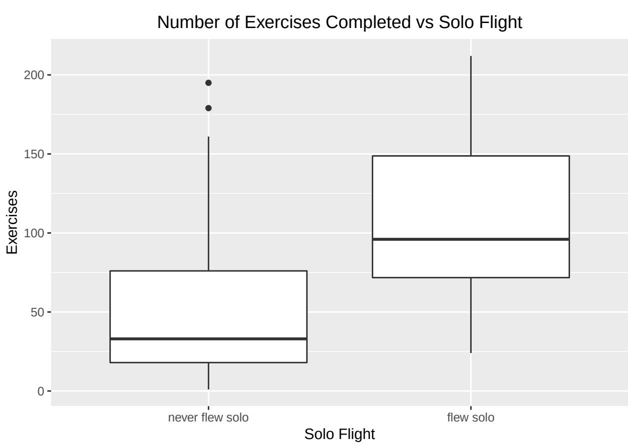


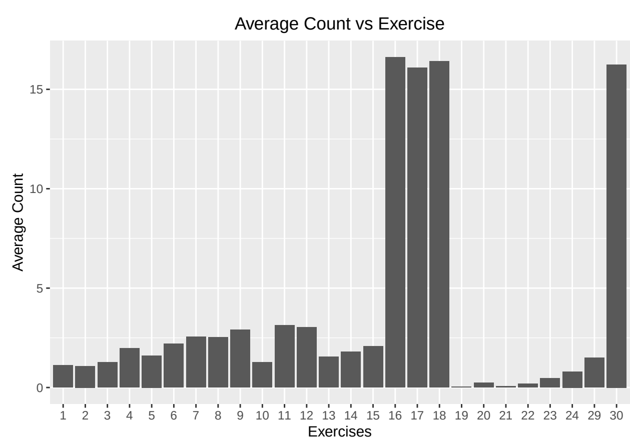


Average yearly hours for each instructor is defined as the difference between the first and last session each instructor conducted divided by the apporximate length of a year (365.25 days). We found that the Pearson correlation coefficient was approximately 0.9827651 indicating a very strong correlation. Based on this, we concluded that the number of exercises students completed under each instructor per hour of work is a strong measure of efficiency. We believe that ideally instructors should be getting a number of hours directly proportional to their efficiency. Thus, we generated the second graph above to determine a roughly ideal distribution of hours for each instructor. Instructors whose points lie above the line are, by our model, being given too many hours relative to average efficiency and instructors below the line are being given too few hours relative to their efficiency. To maximize profits, instructors should be training students as fast as possible during the times of high demand. While there is a minimum hours limit, it is reasonable to believe that instructors that complete more exercises in a set time will have students perform better. Thus, their students will have a better chance of passing the flight exam the first time.'

Student Success

To establish that flying solo can be considered to be a metric for students' progress we tested if there is a difference in the average number of exercises completed by students that flew solo and the number of exercises completed by students that never flew solo. Only exercises coompleted up to and including the first solo flight count for this comparison. The first solo flight is determined to be the first session with solo in training type. We first plotted side-by-side boxplots of the number of exercises completed by both groups of students.





The boxplot suggests that there is a significant difference in the number of exercises completed between both the groups. Hence, we attempted to find the collection of exercises that were typically completed before a student flew solo and graphed the average count of each exercise completed. Additionally, we discovered that each instructor's qualifying collection of exercises to fly solo differed from others' with instructors 3 and 11 requiring lesser frequencies of exercises 16,17,18,30 than the average and 4,6,7,9 requiring higher. Moreover, instructors 10 and 14 do not have a distinguishing qualifying collection with exercise counts for both groups of students are very close to each other (see note 2 in appendix for corresponding tables).

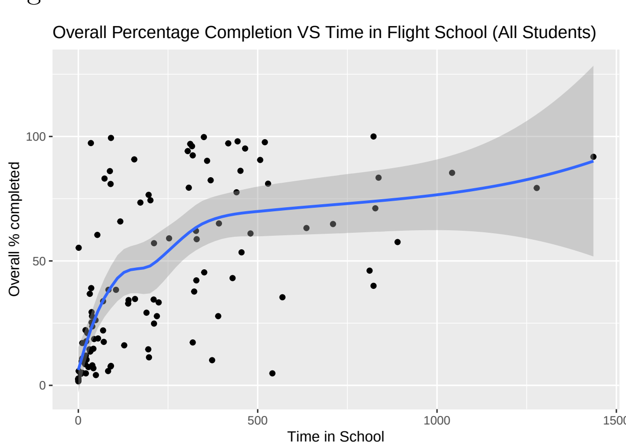
We checked the training data to see if there is a specific order in which exercises should be completed. To determine the order in which the exercises should be completed, we found the session at which at least 70% of students who completed that number of sessions, completed the given exercise.

Table 1: Suggested Order of Exercises

Exercise	Expected_Session	Exercise	Expected_Session	Exercise	Expected_Session
1	1	9	4	17	10
2	1	30	5	13	12
3	1	10	6	15	12
4	1	11	7	29	31
5	1	12	8	19	40
6	2	16	9	22	42
7	3	18	9	24	46
8	3	14	10	23	48

The above table displays the implied order of exercises based on the criteria explained above. This order seems to suggest that while exercises 1 to 5 are important to complete immediately, the later exercises require much more experience, as exercises 19, 22, 23, and 24 were not completed by 70% of students until after 40 or more sessions.

We analyzed how many students met the flight license requirements, and compared their progress based on the amount of time spent in the program. We have analyzed how close each student has come to completing their pilot license requirement, based on each requirement needed to graduate (for exact requirements, see note 3 in appendix). Overall is the weighted mean of each requirement (based on hours), which is used to estimate each students' overall progress percentage towards their flight license.



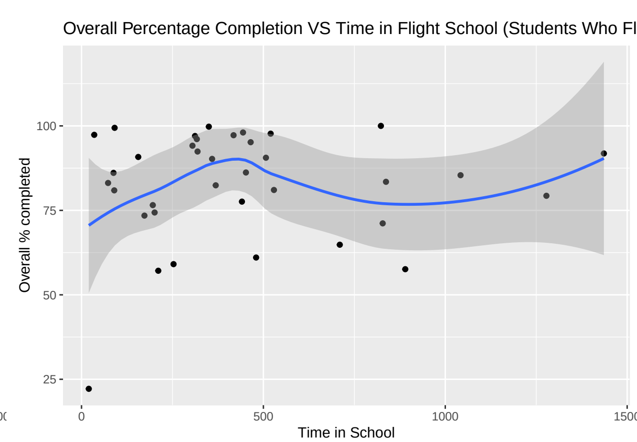


Table 2: Top 30 Students by Course Completion

Student_ID	Overall	Student_ID	Overall	Student_ID	Overall
88	100.00000	41	94.13793	74	83.10345
115	99.77011	42	92.41379	100	82.41379
4	99.42529	120	91.83908	56	81.03448
21	98.04598	77	90.80460	16	80.91954
76	97.70115	38	90.57471	1	79.42529
66	97.35632	87	90.22989	110	79.31034
121	97.24138	40	86.20690	64	77.58621
61	97.01149	6	86.09195	65	76.55172
14	96.09195	117	85.40230	8	74.36782
93	95.17241	53	83.44828	55	73.44828

Note that time in school is defined as the difference (in days) between the first and the last session a student completes. We observed that only Student #88 completed all the requirements for the pilot's license, however multiple students have come very close to completion. Furthermore, from our graph, we see that students that come close to completion, do so in a varying amount of time. Thus we hypothesize that students finish the course at their own pace. We can also see a linear trend in the beginning of the graph which becomes somewhat logarithmic later. This is due to the high number of beginner students with low completion percentages, and the students that complete the course in a longer period of time. Furthermore, when plotting only the students flying solo we see that there is little correlation between time in school and % completion.

Summary

We concluded there are several aspects of the training school's training process that can be adjusted to create a more efficient system. More efficient instructors should be assigned more students, especially during busy periods. This will allow the flight school to maximize the number of students going through the school and thus will increase profits. Additionally, the flight school should aim to train more students in the month of September as wind conditions are similar to the month of August but total air traffic decreases, thus making flying in September safer. A benchmark for student success was determined to be the student's first solo flight. Taking a measure of when 70% of students had completed a specific exercise, we derived a suggested order of exercises. Finally, we concluded that there is currently no set pace for completing the flight course as students with varying course completion have been in the course for inconsistent periods of time.

Appendix

Note 1: We conducted a hypothesis test to validate our conclusion from the boxplot of mean exercises completed between students that flew solo and those that did not at 5% significance level.

$H_0 : \mu_{solo} - \mu_{non_solo} = 0$ vs $H_A : \mu_{solo} - \mu_{non_solo} \neq 0$

Z = -5.7039, p-value < 1e-04

We see that the P-value is indeed very strong (<0.0001). And at 5% significance level we reject the null hypothesis and accept the alternate hypothesis. Thereby, we can conclude that a solo flight is a strong metric for students' progress as it requires a certain level of prior experience and skill.

Note 2: Instructor vs Avg Exercise Count of students before first solo flight

Instructor_ID	16	17	18	30
3	14.50000	14.00000	14.50000	27.50000
4	20.60000	19.40000	20.00000	17.30000
6	17.00000	17.16667	16.83333	17.83333
7	24.75000	24.75000	25.00000	19.00000
9	16.33333	16.33333	16.66667	16.66667
10	5.00000	5.00000	5.00000	9.00000
11	11.50000	11.50000	11.50000	7.50000
12	15.00000	12.66667	14.33333	20.00000
14	8.00000	8.00000	8.00000	8.00000

Instructor vs Avg Exercise Count of students who never flew solo

Instructor_ID	16	17	18	30
1	12.0000000	9.166667	11.666667	5.8333333
3	6.4000000	5.400000	5.600000	12.0000000
4	3.1666667	2.666667	3.083333	3.1666667
6	5.7500000	5.250000	5.500000	7.3750000
7	6.4000000	3.400000	6.200000	8.2000000
9	5.9090909	5.909091	6.000000	5.8181818
10	5.4000000	5.100000	5.200000	5.1000000
11	0.9285714	1.000000	1.000000	0.9285714
12	4.6000000	3.400000	4.200000	9.8000000
14	4.5000000	3.750000	3.500000	3.5000000
15	2.0000000	2.000000	2.000000	1.5000000
17	14.0000000	14.000000	14.000000	24.0000000

Note 3: The exact training time requirements to complete the flight course are listed as follows:

- 45 Total Flight Hours
- 17 Total Dual Flight Hours
- 3 Total Dual Cross Country Hours
- 5 Solo Cross Country Hours
- 5 Instrument Hours

The following pdf contains exact specifications for training requirements:

https://durhamflightcentre.com/wp-content/uploads/2018/09/PPLredbird_Dec15.pdf