Data Analytics Engineering

Spring 2023

United States Aerial Firefighting Capabilities & Operations: Dashboard Reporting and Predictive Analytics

DAEN 690

Project Report

Nikhil Gudupalli
Sai Aishwarya Sakamuri
Spencer Marlen-Starr
Suresh Yeluri
Tanishq Ravi

About the Cover

Dr. Isaac Gang is an Associate Professor at the George Mason University College of Engineering and Computing, Volgenau School of Engineering, MS Data Analytics Engineering (DAEN) program.

He joined the DAEN faculty in the Fall of 2020 from Texas A&M University-Commerce (TAMUC) where he served as an Assistant Professor of Computer Science as well as the department's Outreach Coordinator. Before coming to TAMUC, Dr. Gang was an Assistant Professor of Computer Science and Engineering at the University of Mary Hardin-Baylor (UMHB) and an Adjunct Professor of Computer Science at the University of Southern Mississippi's School of Computing before joining UMHB.

Dr. Gang is a former DOE grant winner, former President and Board Member of the Association of Computer Educators in Texas (ACET), Industry Advisory Board (IAB) Coordinator, and the Director of CS For All.

His current and primary teaching responsibilities at Mason largely involves Data Analytics Engineering graduate courses along with a mix of CS and AIT graduate courses. He is an affiliate faculty member of GMU's C^{4}I & Cyber Center.

Dr. Gang's primary research agenda involves Big Data Analytics (emphasis on data bias and data governance), Cyber Security (ransomware, steganography, and cyberbullying), and Image/Signal Processing.

This page intentionally left blank.

Abstract

Due to both increasing urban sprawl and the effects of anthropogenic global warming, the United States is expected to see an ever-growing number of wildfires every year within its borders over the remainder of the 21st century. As this occurs, the US's aerial firefighting capabilities specifically are becoming more and more of an essential component of the comprehensive resources at its disposal to fight, contain, mitigate, and control wildfires because of the pace of innovation in aviation, both in cost reduction and increased capabilities in aerial firefighting equipment and vehicles far outpaces that of traditional ground-based firefighting equipment and vehicles in our modern age. Currently, the US Forest Service, the Bureau of Land Management, CalFire, etc., are doing their best to manage and coordinate their respective fleets of aerial firefighting aircraft efficiently and effectively. However, this is not feasible to achieve without access to up-to-date country-wide information. For example, in 2019, a severe wildfire in the Amazon rainforest highlighted the importance of aerial firefighting and emphasized the unpreparedness of officials for such an event. During this crisis, Jeff Berry, Director of Business Development Conair, pointed out that there is an immediate need to understand the aerial fire-re-fighting resources such as runways information, availability of aircraft, especially the ones that are under CWN (Call-When-Needed) basis, water resources in the proximity area, etc. The purpose of this report is to produce a working prototype for a real-time dashboard suited for just such a task, namely, aggregating all the data and metrics that high-level stakeholders and decision-makers in Washington or elsewhere would require to be able to manage all our aerial firefighting resources at scale nationwide effectively.

Report

Section 1: Problem Definition

1.1 Background

Wildfires are unplanned fires that begin outdoors, including lightning-caused fires, unauthorized human-caused fires, and escaped prescribed fire projects (Congressional Research Service, 2021). Due to environmental, ecological, latitudinal, and more human-centric factors, the United States is subject to many wildfires yearly. While there are many different areas throughout the contiguous lower 48 States where wildfires can and do happen from time to time, the problem is by no means uniform, with each of those states at equal risk over the coming years. Even though the United States is an enormous country geographically speaking (4th largest in the world by land mass), nearly half of its wildfires every year occur in just one of its 50 states, California. According to Verisk Analytics' most recent State Wildfire Risk Report for California (FireLine State Risk Report-California, 2021), there were over 4 million acres that burned due to wildfires in 2020 just in California 2020, which was well over half of the total land area burned due to wildfires across America that year. Furthermore, looking towards the future, they deem roughly 15% of California's housing stock as facing “High and Extreme” risks from wildfires! Because there are around 14 million houses in California, that puts over 2 million in their highest risk category in just that one state.

Continuing urban sprawl due to suburbanization and anthropogenic climate change are two factors that, in turn are increasing the susceptibility to wildfires of many large swaths of the United States over time. According to the article “U.S. Wildfire Statistics 2022” (Sleight and Kempken 2022), currently, over 4 million homes are at substantial risk from wildfires, and to make matters worse, there was a 17% increase from 2019 to 2021 in U.S. wildfires on top of a whopping 223% increase since 1983! The top 3 years in terms of how much damage caused by these wildfires in monetary estimates have all been recent; the third worst was 2020, a year in which wildfires in the U.S. caused $16.5 billion in damages, the second worst on record was 2018, which saw $22 billion in damages, and the very next year, 2018, being the highest at an astonishing $24 billion. And the same story is a dreary picture painted by the official Wildfire Statistics fact sheet and pamphlet put out by the Congressional Research Service as of its most recent update in

March of 2023, from 2013 to 2022, there were an average of 61,410 wildfires annually and 7.2 million acres of American land impacted by these annually. In 2022, 68,988 wildfires burned 7.6 million acres. Unfortunately, thus far, the US has dealt with the increasing problem of wildfires in an uncoordinated manner, handling them using several different government agencies and departments, and on top of that, there is further variation between managing wildfires at the federal, state, and local levels. Furthermore, all of this was done without a mandate for this diverse set of governmental bodies to share and aggregate their data with one another.

The magnitude and urgency of this threat point to the need to coordinate, organize, and data-driven management, as well as mitigation, and prevention tactics for dealing with this looming threat by the United States government. Clearly, as this problem continues to grow year after year, the current ad hoc approach, which is the status quo will only be able to keep up with the task at hand if it is more scalable. This research report is specifically focusing the United States' aerial firefighting capabilities, an essential subcomponent of the United States' overall portfolio of tools and resources for fighting wildfires which have been gaining in its importance over time for the previous few decades and is likely to continue to do so in the future. That is why, according to a lengthy report commissioned by the European Policy Evaluation Consortium, "Efficient and effective management of Aerial firefighting resources is needed if the twin challenges of growing vulnerability to wildfires and heightened pressure on public finances are to be managed successfully" (GHK Consulting 2010).

Currently, there are many different services. Aerial firefighting can provide and many roles it can play. According to "Fire Aviation Guidelines," a report commissioned by the United Nations to outline a common set of suggested guidelines intended to help countries around the world better manage their fleets of firefighting aircraft, the different roles these aircraft can play include: directly or indirectly attacking a fire by dropping water or fire suppressants and retardants, for a range of tactical purposes; dropping incendiaries to ignite managed fires; transporting and delivering firefighters and equipment to the foreground; providing command and communication platforms; providing fire detection and reconnaissance services; enabling sophisticated mapping and intelligence gathering to assist the management of fire and to provide high-quality information to threatened communities; delivering warnings or evacuation orders to communities; supporting fire prevention and enforcement; and conducting fire damage assessments (International Fire Aviation Working Group 2014).

When it comes to the US Forest Service over the last few years, to get a sense of the scale of the Aerial firefighting resources they have at their disposal, we turn to their official annual reports. In their report for fiscal year 2021 (2021 Aviation Annual Report), "449 contracted aircraft and 25 agency-owned aircraft were utilized to meet agency missions." Meanwhile, for the prior year (2020 Aviation Annual Report), "458 contracted aircraft and 24 Agency-owned aircraft were utilized to meet Agency missions." These annual reports designate all aircraft owned, managed, or used by the United States Forest Service into one of the 6 categories: fixed-wing, helicopter, airtanker, scooper, Modular Airborne Fire Fighting Systems (MAFFS), and unmanned/uncrewed aircraft systems (UAS).

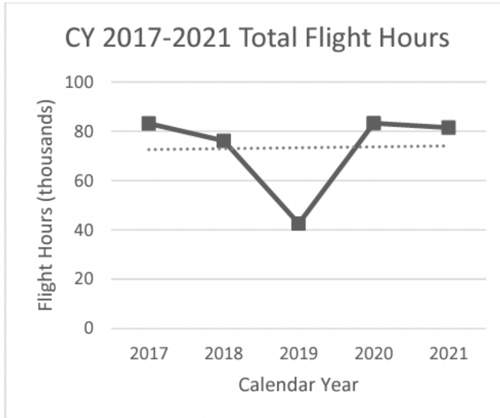
Focusing on just their fleet of planes (fixed-wing aircraft), they group all the different makes and models they use by these 6 categories (Planes | US Forest Service): First, Single Engine Airtankers (SEATs) can deliver up to 800 gallons of fire retardant to support firefighters on the ground. These small airplanes can reload and operate in areas where larger air tankers cannot; airplane models used by the US Forest Service considered as SEARs: Air Tractor AT-802. Second, Large Airtankers (LATs) can deliver from 2,000 to 4,000 gallons of fire retardant to support firefighters on the ground; airplane models used by the US Forest Service considered as LATs: P2V, HC-130H, BAe-146, MD-87, C-130Q, RJ85, C-130 H & J equipped with Modular Airborne Fire Fighting Systems (MAFFS). Third, Very Large Airtankers (VLATs) can deliver over 8,000 gallons of fire retardant to support firefighters on the ground; airplane models used by the US Forest Service are considered as VLATs: DC-10. Fourth, Water Scoopers are amphibious aircraft that skim the surface of a water body and scoop water into an onboard tank, and then drop it on a fire; airplane models used by the US Forest Service considered as Water Scoopers: Bombardier CL-415 and Air Tractor Fire Boss. Fifth, Smokejumper aircraft deliver smokejumpers and cargo by parachute for initial attack and extended support of wildland fires. Each aircraft can carry eight to ten Smokejumpers and their initial supply of gear; aircraft makes and/or models deemed smokejumpers: DeHavilland DH-6 300 series Twin Otter, Shorts Sherpa C-23A and SD3-60, Dornier 228, and CASA 212. And finally, Sixth, the newest addition to the Forest Service's Aerial firefighting portfolio, Unmanned Aircraft Systems (UAS). UASs are hoped and believed to have great potential for use in fighting wildfires and other aspects of proper and wise natural resource management. In contrast, unauthorized public UAS flights over or near wildfires threaten the safety of aerial and ground firefighters, and users are encouraged to "know before you fly."

The Forest Service's Modular Airborne Firefighting Systems capabilities (the 3rd of the 6 categories of firefighting aircraft above) require more depth because they are increasing in their importance and could become key in the future. According to the page dedicated to these resources in their arsenal (Modular Airborne Fire Fighting Systems (MAFFS), 2017), "MAFFS are portable fire-retardant delivery systems that can be inserted into military C-130 aircraft without major structural modifications to convert them into (extant military) air tankers when needed." The way this collaborative effort between the Pentagon and the Forestry Service is that Forest Service owns the MAFFS equipment and supplies the fire retardant. At the same time, the DoD provides the C-130 H and J model aircraft, flight crews, and maintenance and support personnel to fly the missions.

	Number of Aircraft
Helicopters	
Exclusive Use (EXU)	87
Call When Needed (CWN)	204
Agency Owned – WCF	3
Airtankers	
Next Generation – EXU	24
Next Generation – CWN	6
MAFFS	8
Multi-Engine Water Scoopers	
Call When Needed (CWN)	8
Fixed-Wing	
Aerial Supervision Module / Leadplane (Lease)	17
Light Fixed-Wing – EXU	29
Light Fixed-Wing – CWN	92
Smokejumper Aircraft – EXU	8
Smokejumper Aircraft – CWN	0
Large Transport – EXU	2
Agency Owned Fixed-Wing – WCF	22

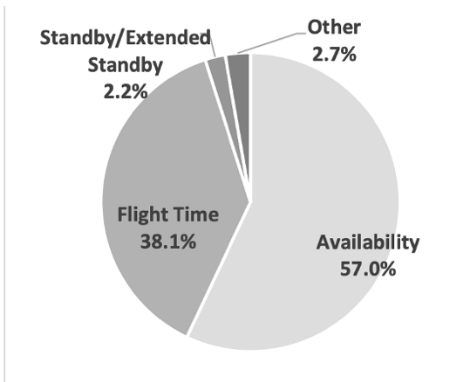
This includes aircraft owned by government agencies themselves and aircraft these agencies contract for as well. The fixed-wing category contains large transport jet, lead planes, smoke jumper aircraft, and the other fixed wing aircraft operating to deliver the fire suppressant. The airtanker category contains all the flights associated with all fixed-wing aircraft delivering chemical suppressant payloads to help control and contain wildfires. The helicopter category contains rotor wing aircraft irrespective of the mission they are being used for. Finally, the scoopers are fixed-wing water scooping aircraft used in fire suppression.

Calendar Year	Flight Hours
2017	83,184
2018	76,230
2019	42,570
2020	83,324
2021	81,487
5-Year Average	73,359



In the calendar year 2021, agency and contract aircraft flew 81,487 hours, which was more than the 5-year average and only 2.2% less than the previous high in recent years, which occurred in 2020.

Pay Code Description	Total Costs
Availability	$428,377,330
Flight Time	$286,261,440
Standby/Extended Standby	$16,454,132
Other	$20,404,097
Total	$751,497,000



1.2 Problem Space

This project aims to create an easily usable, comprehensive dashboard with real-time data that keeps track of all flight operations and aerial resources at the American government's disposal for fighting these increasingly dangerous and erratic yet fundamentally unpredictable wildfires. This real-time dashboard can then be utilized by a high-level decision-making bureaucrat in Washington DC (e.g. Randy Moore, the current chief of the US Forest Service) who has been tasked with making big-picture decisions regarding overall US wildfire fighting policy, approaches, and procedures now, and possible plans. In addition, we hope to create a dashboard useful enough to improve our logistics and capabilities in this country to fight, mitigate, and contain wildfires more safely and effectively using its Aerial firefighting resources.

1.3 Research

Our research on the existing peer-reviewed literature and official documentation put out by the US Forest Service on Aerial Fire Fighting, Predictive Analysis, and practical Real-Time Dashboards has aimed to improve our understanding of how data and analytics can be used to support more effective, efficient, and reliable firefighting operations. Researchers from different disciplines have explored different methods for predicting fire behavior and spread, including classical statistical modeling (multiple regression analysis), machine learning algorithms, and data visualization techniques. These have been undertaken by ecologists, climate scientists, and by data scientists and statisticians. In the process of carrying out this research, they have studied how data from a vast array of diverse sources, such as weather forecasts and data, satellite imagery, data on historical fire impacts and trajectories, case studies, samples taken by ecologists regarding how dry trees in different forests are and, even more importantly, changes in the 'deadwood volume'[1] of individual forests over time that can be analyzed. Taken together, the insights gleaned from all of this can, in principle, be integrated coherently to support better firefighting decision-making.

Studies have also investigated the use of predictive analysis and dashboards in real-world firefighting scenarios, evaluating their impact on resource allocation and fire suppression strategies. Results have shown that using predictive analytics combined with well-designed dashboards leveraging real-time (or quasi real-time) data can improve situational awareness, increase the efficiency of firefighting operations, and enhance decision-making quality.

Overall, research in this area has helped develop more effective and efficient aerial firefighting systems and has shown the potential of data and analytics to support firefighting efforts. Ongoing research is likely to continue exploring new methods for predicting fire behavior, improving data integration and analysis, and further refining predictive analysis and dashboards in firefighting operations.

[1] Deadwood is dead or rotting branches, stumps, or entire fallen trees scattered throughout a forest's floor.

1.4 Solution Space

The solution space for the Aerial Fire Fighting Flights Dashboard and Predictive Analysis refers to the area of focus or the set of problems and challenges that the solution addresses. In this case, the solution space for this project focuses on improving the efficiency and effectiveness of firefighting operations that involve fire-fighting aircraft. This solution provides a comprehensive platform for managing and analyzing fire-fighting processes, including real-time monitoring and optimization of resource allocation. The solution space encompasses a wide range of stakeholders, including fire managers, incident commanders, firefighters, support team members, and decision-makers, who can all benefit from the real-time information, insights, and guidance the solution provides. By addressing the key challenges and goals of firefighting operations, the Aerial Fire Fighting Flights Dashboard and Predictive Analysis solution aims to enhance fire suppression efforts' safety, efficiency, and effectiveness.

1.5 Project Objectives

Our first objective is to learn how to create a comprehensive, real-time dashboard.

The second objective of this project, after creating the initial version of the dashboard, is to enable it to handle updates from streaming data sources occurring at least once per day.

Our final objective for this project is to include at least one graph of a forecast of some critical aspect of the national management of aerial firefighting resources, such as, when certain older planes will need to be replaced as the entire fleet continues to depreciate over time (but this is just an example, it could be something else).

And an additional instrumental objective all of us must learn in the process of completing this project, and quickly for that matter, is for all of us to master the art of using, implementing, and adhering to the Agile Scrum Methodology on YouTrack professionally. This will be instrumental in our ability to achieve the three of the mentioned goals.

1.6 Primary User Stories

1. As a user, I want to keep track of the efficiency and cost of aerial firefighting operations to ensure that the resources for aerial firefighting are used in accordance with it. The information such as flight hours and the cost incurred per flight hour, the volume of chemicals used to suppress the fire, and the impact of the effort will be helpful in making well-informed and sound decisions about the allocation of resources and planning for future aerial firefighting operations.

2. As a user, I need access to a database containing comprehensive information on all aircraft designated and equipped for aerial firefighting operations. Information such as the types of aircraft, the registration number, the model, the year of manufacture, and the manufacturer's name if I am to manage and track the aircraft information in the fleet efficiently. To have access to the most up-to-date fleet information, the information should be updated in real-time. This data ensures the operations' safety and efficacy.

3. As a user, I'd like access to a comprehensive airport information database so that I can easily manage and track the information of all the airports in my network. This database should contain information such as the airport's name, location, runway lengths, elevation, and other relevant information. Furthermore, I'd like to be able to update this information in real-time so that I always have the most recent information about my airports. This information is critical for ensuring my aviation operations' safety, efficiency, and compliance.

4. As a user, I want to create a real-time dashboard that integrates all pertinent information about firefighting operations and enables me to make informed decisions in real-time. This platform should include data from various sources, including real-time maps, resource allocation, and aircraft data. I intend to use historical data and predictive analytics to identify and allocate resources based on the risks involved. This platform will help me manage and coordinate aerial firefighting operations effectively.

5. As a user, I want to be able to plan and prepare for aerial firefighting operations in advance so that I can respond to wildfires as effectively as possible. This system should allow me to plan routes, and prepare equipment, and personnel. I want to be able to predict potential risks and make informed decisions using predictive analytics. This will greatly aid in the well-organized response to wildfires.

1.7 Product Vision

For: Our Project is for decision-makers in flight operations and aerial firefighting such as civil servants in the federal wildfire system, flight coordinators, air operations supervisors, and fire management officers.

Who: who is responsible for monitoring, planning, and executing flight operations for aerial firefighting.

The: Real-time dashboard and predictive analytics analyze

Is a: Historical and Real time data on flight operations and firefighting efforts.

That: Identifies trends, patterns, potential risks, and benefits decision-makers in making informed decisions on flight operations and aerial firefighting.

Our product: assists in flight operation preparedness and increases efficiency in aerial firefighting operations.

Scenario #1

The project would benefit government agencies and other stakeholders, as it would provide them with the crucial information and insights required to formulate and implement regulations and policies related to aerial firefighting and flight operations.

One of the primary benefits of the final dashboard for government agencies would be the ability to monitor and analyze historical and real-time data on flight operations and firefighting efforts. The data can be used in identifying trends, patterns, and potential risks and benefit government agencies in making informed decisions on flight operations and aerial firefighting. The project helps in identifying fire zones that are associated with increased risk and helps in decision making to mitigate the risks by increasing the firefighting resources in the area. The project's dashboard and predictive analytics solution can be used to track the performance of flight operations and aerial firefighting, as well as to identify any areas where compliance may be an issue and help the agencies in increasing the efficiency of aerial firefighting operations. The project assist decision makers in flight operation preparedness and be better prepared for changing needs and requirements.

Scenario #2

The project would help gather the characteristics of each airport location in the United States of America. Then, in a time of need, effective decision-making can be made regarding the safe landing of the aircraft, as the data for the runway information would be provided. Not only the runway information but also aims to include the navigation equipment for each airport facility. With the information available at a click of a button, quick and sound decisions can be made when the fire-fighting aircraft in action need immediate safe landing and refueling, thus, ensuring fire-fighting aircraft safety.

Scenario #3

The project would benefit owners and operators of aerial firefighting companies and airport managers. The dashboard and predictive analytics solutions provide valuable insights to the owners and operators of aerial firefighting companies about the performance of their operations. This helps in increasing the efficiency and helps in identifying crucial areas where improvement is required. The dashboard helps optimize resource allocation, improve safety, and reduce costs. For airport managers, the project helps manage the security and safety of airport operations and improves coordination between emergency response organizations in their firefighting efforts. The predictive analytics solution enables airport managers to be better prepared for future emergency situations. Overall, the project will provide valuable resources for airport managers and owners of aerial firefighting companies to optimize their operations, and improves safety and preparedness for future incidents.

Section 2: Datasets

2.1 Overview

We explored many possible sources to obtain data on the Tail Id numbers for all Aerial firefighting aircraft (both fixed wing and rotary alike), airport ID numbers for each airport these aircraft took off from or landed at, the different types (their make & model) of aircraft used for Aerial firefighting, number of flight operations per day and their average duration, the number of take-offs each day at each airport, the number of runways at each airport and the weight bearing capacity of each of those runways, and the ratio of the total number of operational firefighting aircraft within a state to it the area of that state (this is a very simple arithmetic operation to perform and is done quite often for many different things, for instance, when this same ratio is calculated for people living in a state instead, it is known as the population density of that state). But in the end, we only needed to use three of them: the United States Federal Aviation Administration's (FAA) online Aircraft Inquiry Registry/Database, the FAA Airport Database, and FlightAware's Flight Tracking Database.

Because the FAA's Aircraft Inquiry website is a .gov and because the FAA holds jurisdiction over all flights, aircraft, and airlines within the United States, the data it has on many aspects of flight operations within the United States must be reported to them by the current owners of all aircraft actively being used for flight operations every year. In this sense, the data from this source avoids many of the most common problems encountered when conducting most empirical research and/or analytics projects in the real world. For instance, reporting the data about an aircraft's N-Number, make/model, serial number, engine type, state, manufacturer, document index, flight ID, airport of origin, destination, time of day, flight

Duration, etc. is mandated rather than requested, we are not dealing with a sample of flight operations; we are dealing with the entire population.

According to the 'About' section of their website, FlightAware is a "digital aviation company and operates the world's largest flight tracking and data platform." FlightAware's business model revolves around providing real-time, historical, and predictive flight tracking data, aviation data, and predicted ETAs to airlines, airport operators, and software developers. They provide all these services for both private and commercial flights. In addition, flightaware.com includes informational and organizational services for pilots, such as flight planning, aviation news, photos, and even an aviation discussion forum.

2.2 Data Context

The FAA airport dataset is an Excel Workbook, having 4 sheets of detailed data, is unique in the sense of being non-redundant and has the potential to be very helpful for any federal agencies involved in aerial firefighting in the United States within the context of identifying crucial airports and landing sites for the aircraft used in aerial firefighting operations. This is a complex dataset; the airport_numbers field contains identifiers for various airports. The range of data starts with sheet 1, which contains airport information such as Loc Id (Location Id) and ICAO Id (International Civil Aviation Organization Id), the facility type, and geographical information such as latitude and longitude coordinates, state, address, etc. And also, most importantly, the facility type is categorized as airport, heliport, sea base plane, glider-port, balloon port etc. Sheet 2, on the other hand, dives deeper into the runway information for each facility. The runway id, the dimensions for each runway, the runways' elevation, the runways' weight-bearing capacity, and other navigation equipment such as edge light intensity, wind indicators, etc. Remarks in sheet 3 is the only distinctive column, as the others are just repetitive information such as ICAO ID, the State in which the airport is present and remarks which give recently observed changes in the particular airports. The last sheet provides the operation hours for each airport's control tower.

The FAA N Registry contains information about aircraft registrations, such as aircraft ownership, model, and type. It also includes manufacturing information, engine type, and other airport-related attributes. The flight numbers are unique identifying numbers associated to each aircraft. The dataset contains information about the aircraft's registered owner, the manufacturer's name, the expiry date of the aircraft's registration, and aircraft type, such as helicopter or fixed wing. It also contains information about the manufacturer of every aircraft's engine, such as the model of the engine and crucial information about registration. This contains a wealth of information that could be helpful for stakeholders in both the planning and execution of aerial firefighting operations.

The information gathered from "Flightaware" gives real-time data of flight operations for each tail number (aircraft). The airport from which the flight has taken off and landed, travel duration in minutes, and the local time at each airport facility.

2.3 Field Descriptions

Field Name	Description	Data Type	Source
flight_numbers	The nationality and registration markings of U.S. registered aircraft.	Float64	Original
reg_owner_name	The first registrant's name which appears on the Application for Registration of aircraft	String	Added
model	The type of the aircraft model	Float 64	Added
type_aircraft	The classification of an aircraft based on construction and method of operation (glider, balloon, blimp/dirigible, fixed wing, or rotorcraft).	String	Added
manufacturing_name	An aircraft having multiple owners who have purchased an expressed name of the aircraft.	String	Added
exp_date	The date a reserved N-Number was renewed. Reserved N-Numbers must generally be renewed annually.	Float64	Added
eng_type	The type of engine(s) installed on the aircraft (none, reciprocating, turbo prop, turbo shaft, turbo jet, turbo fan, or ramjet).	Float64	Added

date_change_authorized	Date the Aircraft Registration Branch issued the Assignment of Special Registration Marks, AC Form 8050-64.	Float64	Added
type_registration	The type of Registration	String	Added
dealer	A party engaged in the business of manufacturing, distributing, or selling aircraft who has been issued a Dealer's Aircraft Registration Certificate, AC Form 8050-6. Aircraft registered in the name of a Dealer will reflect "None" in the certificate issue date field.	String	Added
mode_S_code_16	Transponder address code assigned to each U.S. Registration Number as part of the Mode Select Beacon System (Mode S). Sometimes referred to as the ICAO address code	Float64	Added
fractional_owner	An aircraft having multiple owners who have purchased an expressed percentage of the aircraft.	String	Added
street	The street address which appears on the Application for Registration, AC Form 8050-1	Float 64	Added
County	The county name which appears on the Application for Registration, AC Form 8050-1	String	Added
Country	The country name which appears on the Application for Registration, AC Form 8050-1	String	Added
state	The state which appears on the Application for Registration, AC Form 8050-1	String	Added
Zipcode	The postal Zip Code which appears on the Application for Registration, AC Form 8050-1	Float64	Added
type_certificate_data_sheet	If registered based on a dealer's certificate, "None" will appear in this field	Float64	Added
engine_manufacturer	The name of the Engine manufacturer	String	Added
engine_model	A code assigned to the aircraft engine_model	Float64	Added
Awdate	Date of Airworthiness	Float64	Added
type_certificate_holder	The type of certification for holder	String	Added

Classification	The Airworthiness classification/category according to the latest Application for Airworthiness Certificate, FAA Form 8130-6.	Float64	Added
category	The Airworthiness classification/category according to the latest Application for Airworthiness Certificate, FAA Form 8130-6.	String	Added
exception_code	To check the information contained in this record should be the most current Airworthiness information available in the historical aircraft record or not	String	Added
Status	Aircraft assigned to the manufacturer under the manufacturer's Dealer Certificate	String	Added
serial_number	The unique number assigned by the manufacturer/builder.	Float64	Added
pending_number_change	The Aircraft Registration Branch has issued an Assignment of Special Registration Marks, AC Form 8050-64, authorizing the new (pending) N-Number to be placed on the aircraft.	Float64	Added
certificate_issue_date	Date the Aircraft Registration Branch issued the Certificate of Aircraft Registration, AC Form 8050-3	Float64	Added
mfr_year	Year the aircraft was manufactured based on information shown on the Application for Airworthiness Certificate, FAA Form 8130-6.	Float64	Added
mode_S_code_8	Eight-digit transponder address code assigned to each U.S. Registration Number as part of the Mode Select Beacon System (Mode S). Sometimes referred to as the ICAO address code.	Float64	Added

FLIGHTAWARE.COM

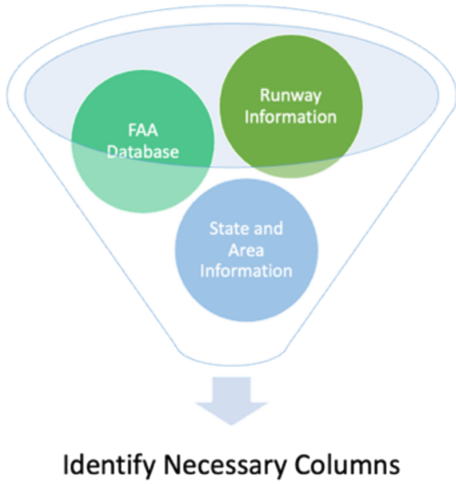
Field Name	Description	Data Type	Source
Date	Date of the flight operation	object	Added
Aircraft	Type of Aircraft	object	Added

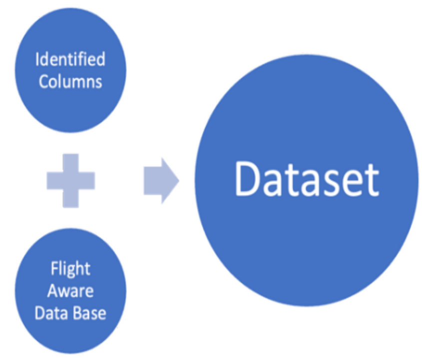
Origin	location of Departure	object	Added
Destination	location of Arrival	object	Added
Departure	Time of Departure	object	Added
Arrival	Time of Arrival	object	Added
Duration	Duration of the flight	object	Added
Tail	Tail no. of the flight	object	Original

2.4 Data Conditioning

After successfully extracted the datasets from the “FAA” databases for aircraft and airport information. There are multiple columns with information about airports and aircraft. The next task would be identifying the necessary attributes/columns to include in the dashboard. As described previously, “All_airport_dataset” is a massive dataset with 4 sheets. Every piece of information we look at is exciting and it is all too tempting initially to want to include far too much of it in the dashboard. However, as a general rule, a lean dashboard that still manages to get its point across is preferable to a dashboard that is so cluttered with minute details that the stakeholders are likely to miss the forest for the trees. Because we seek to avoid overcrowding our dashboard, we organize the information presented into discrete chunks.

And as for FlightAware, with the help of Python programming language, we can scrape this data and store it in an S3 bucket. This can be done using various libraries like BeautifulSoup and Requests. S3 is a cost-effective and scalable storage solution provided by AWS. We can automate this process using a cronjob, a time-based job scheduler in Unix-like operating systems. We can schedule the Python script to run every midnight, updating the data file in the S3 bucket. However, this can lead to the accumulation of duplicate rows over time. To avoid this, we can use Pandas, a data manipulation library in Python, to remove the duplicate rows from the updated data file and get the most recent flight information. This will ensure we always have updated and accurate data in our S3 bucket.



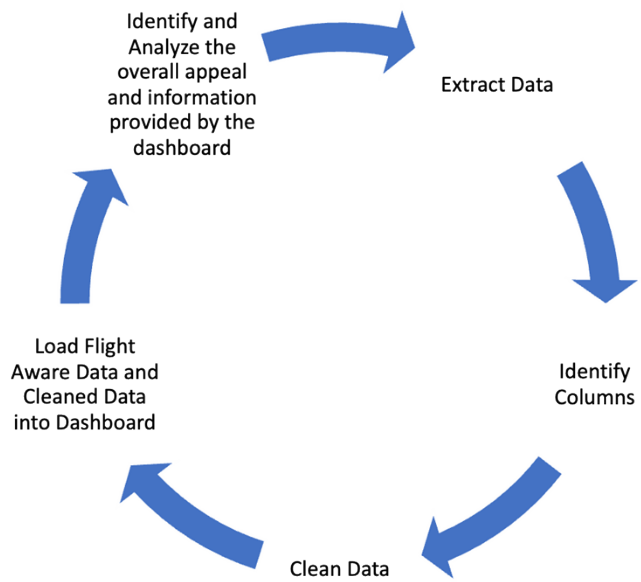


We have decided to partition our Aerial firefighting resources dashboard into three parts or sections. The first part would be the "FlightAware" dataset, which provides real-time information on the current status of ongoing aerial firefighting operations in the air or which took place very recently. The information on these operations provided within that dataset are things like each the current location of the aircraft, which are operational 'right now' so to speak, where they took off from, where they flew to, as well as the local time, date and duration of these flights.

The following section would be aircraft information which would place information about the aircraft such as its registered owner, type of aircraft, whether it is a helicopter or a plane, and using the manufactured year column, we must calculate the age of the aircraft.

Finally, the last section of the dashboard would be runway information. This runway information would constitute navigation aids, and weight-bearing capabilities. Since the number of runways each airport has is necessary for the dashboard, a new column must be inserted that returns such information.

This nominal data must be cleaned by removing null values. The most difficult part would be missing information that is present and available in the FAA databases. So more information would be needed to look appealing to the dashboard.



After finally identifying the required variables or columns, 6 fields were filtered out from the FAA N-Registry Data Set. Which are as follows:

Field Name	Description	Data Type	Source
flight_numbers	The nationality and registration markings of U.S. registered aircraft.	Float64	Original
reg_owner_name	The first registrant's name which appears on the Application for Registration of aircraft	String	Added
model	The type of the aircraft model	Float 64	Added
type_aircraft	The classification of an aircraft based on construction and method of operation (glider, balloon, blimp/dirigible, fixed wing, or rotorcraft).	String	Added

engine_model	A code assigned to the aircraft engine_model.	Float64	Added
mfr_year	Year the aircraft was manufactured based on information shown on the Application for Airworthiness Certificate, FAA Form 8130-6.	Float64	Added

Using the last column, the year in which the aircraft was manufactured,the age of the aircraft was caluculated by subtracting it from the current year.

Age of Aircraft = Current Year – Manufactured Year hence transforming the last column to age of the aircraft.

The FAA airport information was vast, and every piece of information seemed crucial, and narrowing down the selective few was difficult. But keeping the vision of the project in mind and accessing the key factors that would play a crucial role in assisting aircraft in a time of need, a few elements were narrowed down as below. These are type of the facility, latitude and longitude coordinates, navigation equipment, and weight-bearing capacities. A "count" of runways was calculated from runway ids to deduce the number of runways at each airport.

FAA AIRPORT INFORMATION

Field Name	Description	Data Type	Source
Dep_Airport_ID	Airport id of departure airport	Object	Original
ARP_Latitude	Airport Latitude	Object	Original
ARP_Longitude	Airport Longitude	Object	Original
Facility_Type	Type of the Facility	Object	Original
Wind_Indicator	Wind indicator	Object	Original
Runway Id	Id of runway in a facility	Object	Original
Length	Length of the runway	int64	Original
Width	width of the runway	int64	Original
Surface Type Condition	type of the surface type	Object	Original
Edge Light Intensity	intensity of the runway end light	Object	Original

WBC Single	Weight baring capacity of single	float64	Original
WBC Dual	Weight baring capacity of duel	float64	Original
WBC Two Dual	Weight baring capacity of two duel	float64	Original
WBC Tandem Two Dual	Weight baring capacity of Tandem two duel	float64	Original
Base Marking Condition	Base Marking Condition	Object	Original
#Runways	No. of runways in a facility	int64	Original

2.5 Data Quality Assessment

Completeness: The completeness of the datasets related to aerial firefighting is generally good and contains all the necessary information required for the dashboard. Some of the data on certain specific aircraft are incomplete or missing, and the same goes for runways as well, unfortunately. However, the extent of how much data on several of these crucial features appears to be somewhat limited.

Uniqueness: The aerial firefighting datasets contain unique information with most of the entries being distinct. There are some fields that are duplicated and few naming conventions related to aircraft information are inconsistent that influences the usefulness of the data.

Accuracy: The accuracy of the aerial firefighting data is good. Most of the information on the datasets are accurate and reliable. There are not many errors or major discrepancies in the data related to aircraft operations and fuel prices.

Atomicity: The atomicity of the datasets related to aerial firefighting is high, with each entry in the dataset representing information related to firefighting in the United States. There are only a few fields in any of our datasets that are grouped together in such a way that they are difficult to analyze.

Conformity: the overall level of conformity across our datasets is good. The datasets have consistent formatting and data structures across different entries. There are few instances of data recording in various formats with inconsistent units making the analysis part difficult.

Overall, the aerial firefighting datasets are good in quality; however, there are few inconsistencies and limitations that need to be taken care of during the cleaning and analysis.

2.6 Other Data Sources

At the start of this project, we had a lot of things running in our minds to approach the project's end product. We then started scraping all the data that we needed from different websites. We scraped a lot of datasets and did our in-depth research about different kinds of data we needed. Halfway through Sprint 2 we filtered the data we had and decided not to use a few datasets moving ahead in the project.

We scraped data about the fuel prices in different regions of the country. We initially thought that might have an impact on the flight operations cost and could reduce it. Later, we decided not to use it because it doesn't show any big impact on the cost of the operation. Though there was a difference, but it was so minute that we decided not to use it in our analysis.

We gathered data from NPIAS and FAA Navigation aids database. We planned on generating statistics from this data but discovered that this might not be useful. Instead, we gathered airports data and the characteristics of the airports from other data sources. NPIAS data only has data about federally recognized airports but not any helipads. That data might not give full insights for fighting. So, we decided not to use that data.

2.7 Storage Medium

S3 buckets offer a versatile, scalable, and economical alternative for storing and retrieving data in the cloud. S3 is a well-liked option for a wide range of applications and use cases since users can store and retrieve substantial volumes of data from practically anywhere on the web. S3 buckets offer high levels of scalability, durability, and security, and these features—including server-side encryption, access control procedures, and user authentication—ensure that data is shielded from unwanted access.

S3 buckets' versatility is one of its main advantages. S3 may be used for a broad range of applications, including data lakes, content distribution, and archiving. It supports a number of data kinds, such as photos, videos, log files, and backups. S3 also provides a pay-as-you-go pricing mechanism that enables customers to only pay for the storage and bandwidth they actually use, making it an affordable option for storing and retrieving data. S3 buckets are a robust and dependable storage option that can be utilized for a variety of use cases, making them a crucial part of several cloud-based programs and services. (Amazon AWS).

2.8 Storage Security

Storing the aerial firefighting data for the final dashboard in S3 buckets provides high level of storage security. The wide range of features provided by the S3 bucket such as encryption and access controls are highly secure and protects data.

S3 buckets encrypts data through server-side encryption which provides high level security and protects against data breaches. The access control features offered by S3 bucket allows the users who have access to control the data and the actions they perform on the data.

S3 bucket also provides versioning and helps in maintaining a registry on the changes made to the data and provides extra security against deletion and modifications done accidentally.

Storing the aerial firefighting data in S3 bucket provides a highly secure storage solution for our project by protecting data against data breaches and unauthorized access, as implementing encryption measures ensures high level security for our data.

2.9 Storage Costs

The storage cost for using Athena, S3 bucket and Glue job depends on several factors like amount of data stored, access frequency, amount of processing required.

S3 Storage: The storage cost for using S3 bucket depends on the storage capacity used. In general, it cost around $0.023 per GB of storage.

Athena: The storage cost for using Athena depends on the amount of data scanned during queries. For a TB of data scanned it costs around $5.

Glue Jobs: Glue Jobs charge is based on the amount of time they run, and the units of processing power used. A Glue data processing unit costs around $0.44 per hour and 2 is the minimum number of Data processing units required.

Based on the above factors, the storage cost for using S3 bucket, Athena and Glue job for our project is estimated around $250.

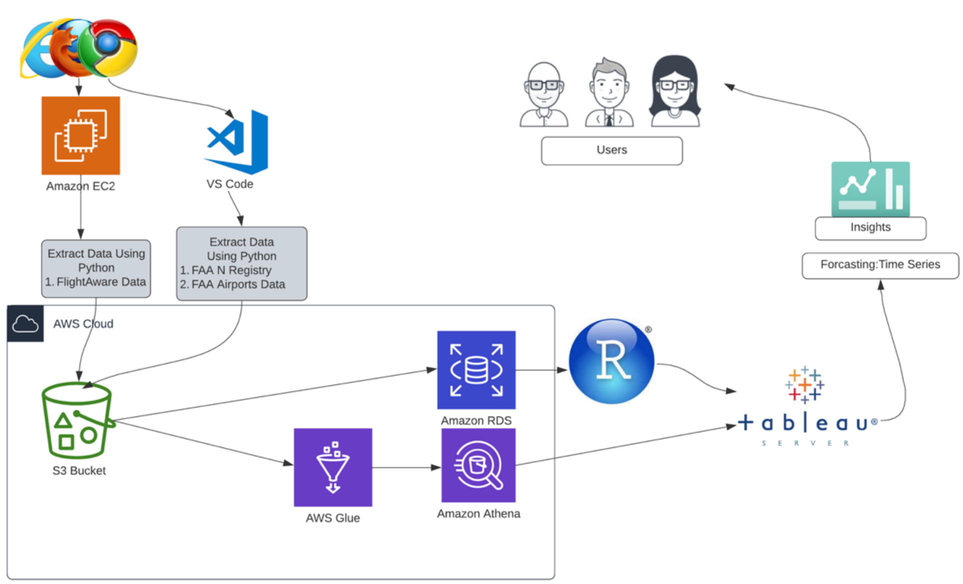
Unfortunately, while the above estimates were conservative and reasonable, we made a mistake involving leaving DB instances open in Amazon RDS because we were under the impression that only EC2 instances left open were what really run up the big-time charges but we were mistaken and this eventually caused our total cumulative costs to absolutely skyrocket to over two thousand dollars. Unfortunately, this is the risk one runs when using cloud services for the first time, we should have kept more of an eye on our costs, and we were at the beginning, but like I said, we thought the only thing we needed to do was to not leave open any large EC2 instances and we were just plain wrong about that.

Section 3: Algorithms & Analysis

3.1 System Architecture

All in all, from top to bottom, in this project, we use 4 distinct tools in the toolbox of the seasoned data analytics professional: Amazon Web Services, MySQL, Python, RStudio, and Tableau. However, all of these are performed on our group's AWS account or through a plugin or connection. MySQL queries and commands are run on our S3 Bucket by way of Amazon RDS and Athena, meanwhile, Tableau is connected to our AWS account using Athena, and the Python code is run using lambda.

Here is a high-level conceptual diagram of the analytic architecture we used for this project:



What was described in the previous paragraph was our goal for our project from halfway through the course, but it did not end up working out that way because we never fully figured out how to get our Access Key ID to connect Amazon Athena to Tableau Server, so we just created the dashboards in Tableau Desktop using data obtained in the aforementioned ways and cleaned again using a combination of Python, RDS via MySQL Workbench, and Amazon Athena. Our time-series analysis towards the end was performed in a combination of Tableau Desktop and RStudio.

3.2 Systems Security

Systems security is critical for any project that involves sensitive data. Data analytics tasks such as cleaning, wrangling, and preprocessing the data before visualizations are performed throughout the various stages of this aerial firefighting project. All the data used should be protected from unauthorized access and modification at each and every step throughout the process. A single Joint Amazon web services (AWS) account is being used by all 5 members of team as the backbone for this project; thus, our security is simply the security provided to us by Amazon automatically in our join account. This is not bad news or as limited as it may sound to the uninitiated with AWS because AWS provides various security measures like access controls, encryption, monitoring, and auditing to ensure the security and privacy of the data stored on its servers.

Access controls are used to protect data from unauthorized access. AWS's control mechanisms regulate access and ensure that only legitimate users with legitimate needs have access to the data. Encryption includes server-side encryption and client-side encryption. These options provided by AWS protect the data at rest and transit from unauthorized access and disclosure.

AWS provides critical security measures such as monitoring and auditing. AWS's monitoring and auditing tools enable the team to monitor and audit the system, ensuring compliance with security and regulatory requirements.

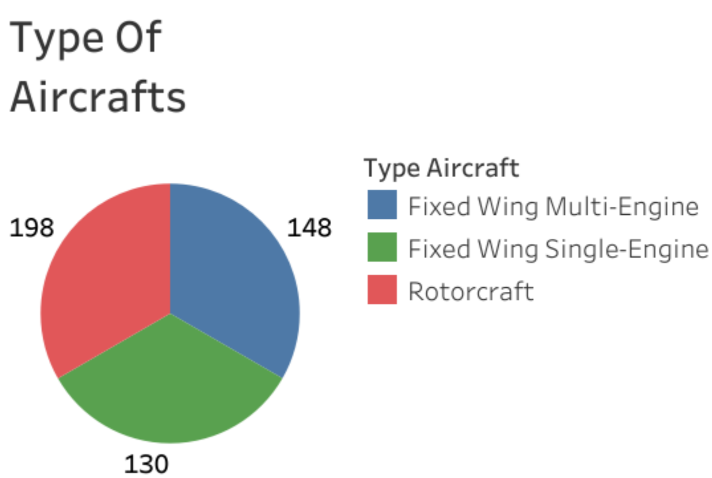
3.3 System Data Flow

Most of the data in the Real-Time Aerial Firefighting operations and resources dashboard we are constructing was obtained by scraping it off webpages using Python. However, this is not the only way we obtained our data. We were also lucky and diligent enough to find a couple of usable and useful datasets which could just be downloaded directly as CSV files. The dashboard itself has 3 different areas of emphasis: 1st, data on airports used to house aircraft used for Aerial firefighting, or airports which they can land on after dispensing their payloads of water or other fire retardants; 2nd, data on all flight operations involved in Aerial firefighting; and 3rd, data on all runways in the United States used by our domestic fleet of Aerial firefighting aircraft, whether the manner in which they are used takes the form of runways firefighting aircraft are taking off from when heading towards wildfires which must be fought or whether they are just a place for these aircraft to land on after returning from their missions, whether that be for the rest of the day, or to refuel and renew their full firefighting capabilities.

Almost all the munging, cleaning, transforming, and preprocessing of the flight operations and airport datasets was done using Python right after it was scrapped. That being said, this was not the case for our data on the airports and runways used by aircraft which performed Aerial firefighting operations. All of the aforementioned standard data analytics operations, which mostly fall under the ETL (Extract Transform Load) category of tasks to be done during a complicated and messy real-world analytics project, were completed using a combination of Amazon's Athena, Glue Crawler, and RDS Services available for us 'on the cloud' via the AWS account which was graciously procured and curated for our team by Dan O'Brien.

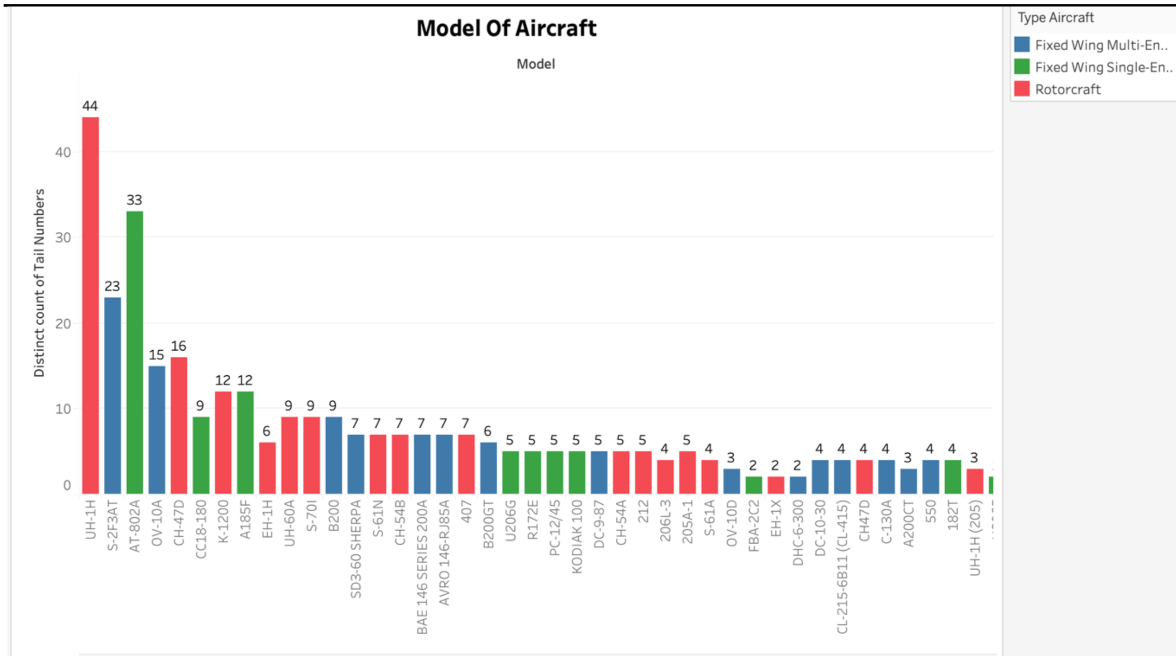
3.4 Algorithms & Analysis

After setting up the data flow, from just the Aircrafts N- Registry database, we were able to start creating some parts of the dashboard. This helps users answer and navigate through the type of aircraft (Fixed Single Engine, Fixed Multi-Engine, and Rotorcraft), their age, and ownership. There are 198 rotorcrafts,130 fixed-single engines, and 148 multi-engine aircraft.

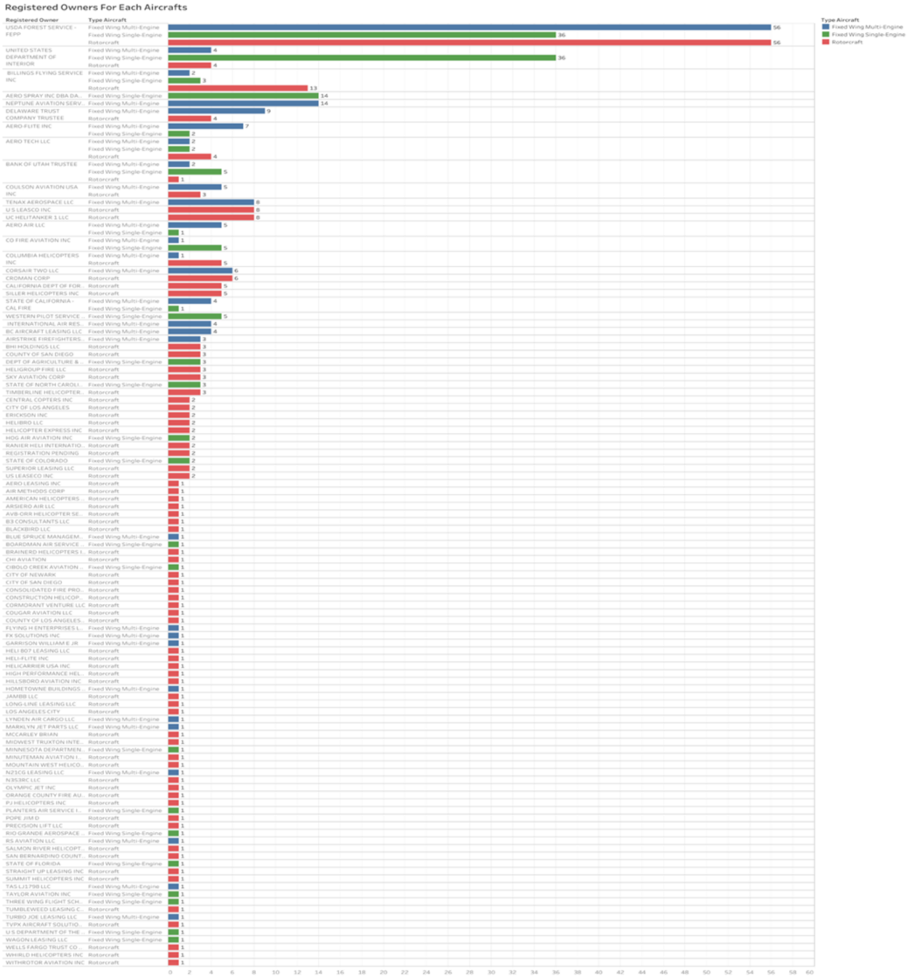


Model Of Aircraft

The model type of aircrafts used in aerial firefighting are displayed in this visualization along with a count of their tail numbers. The list is quite long and includes both single- and multi-engine fixed-wing aircraft in addition to many rotorcrafts. Among the given tail numbers, the maximum Rotorcrafts are of the UH-1H model, which count to 44. There are 23, S-2F3AT fixed wing multi-engine and 33, AT-802A, fixed wing single-engine.



It was discovered that USDA Forest Service FEPP owns more of these aircraft than any other player involved in Aerial Firefighting in the United States. FEPP refers to The Federal Excess Personal Property program. The FEPP program loans these aircraft to State Foresters for wildland and rural firefighting. Most of these aircraft/properties originally belonged to the Department of Defense (DoD), which were later loaned to USDA. This program has been active since 1956. The below graph displays the aircraft ownership information. The USDA Forest Service registered 36 fixed-wing single-engine aircraft, 56 fixed-wing multi-engine aircraft, and 56 rotorcraft.


Registered Owners For Each Aircrafts
Registered Owner
Type Aircraft
Type Aircraft
Fixed Wing Multi-Engine
Fixed Wing Single-Engine
Rotorcraft
Distinct count of Tail Numbers

Sheet 22


Type Aircraft
Aircraft Age
Fixed Wing Multi-Engine
Fixed Wing Single-Engine
Rotorcraft
Distinct count of Tail Numbe..
0
5
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 71 72 81

The visualization displays the age distribution of aircraft. Age ranges for fixed-wing multi-engine aircraft are 5 to 69 years. Most aircraft are 20 to 40 years old. And there are fixed-wing single-engine aircraft that are over 70 years old.

Interestingly when exploring the age of aircraft, N422DE is the oldest aircraft in this dataset which is being used in aerial firefighting and was manufactured in the year 1942. Its current age is 81 years! But it is still being used in Aerial operations by its owner, Cibolo Creek Aviation.

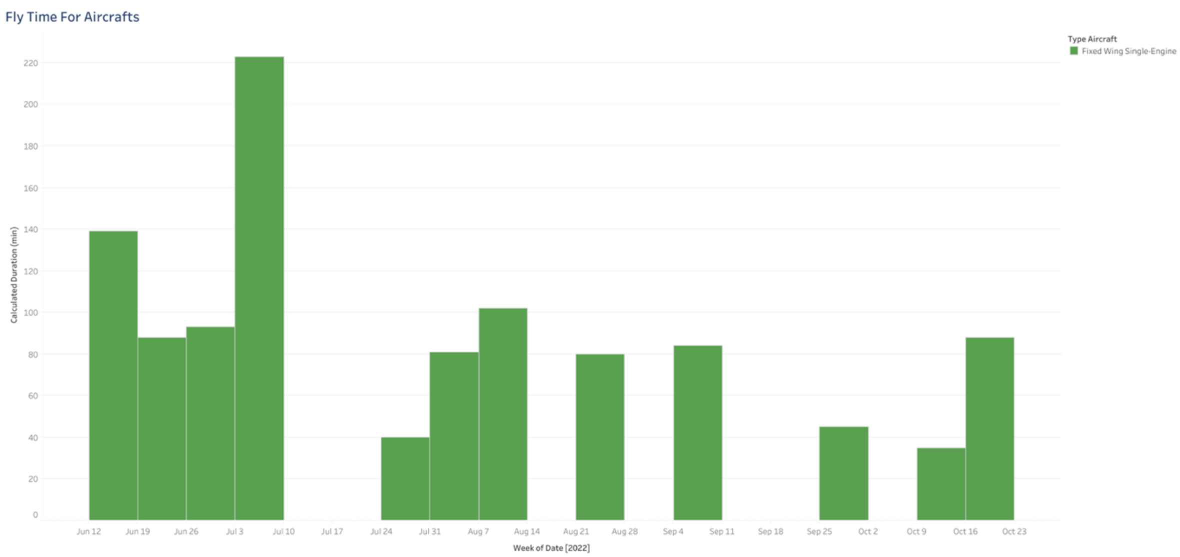
Further landing gear and aircraft models were added to this dashboard part. Therefore, this information needs to be extracted. And data filtering needs to be done as Landing gear information is only relevant to fixed-wing aircraft.

Landing gear configuration

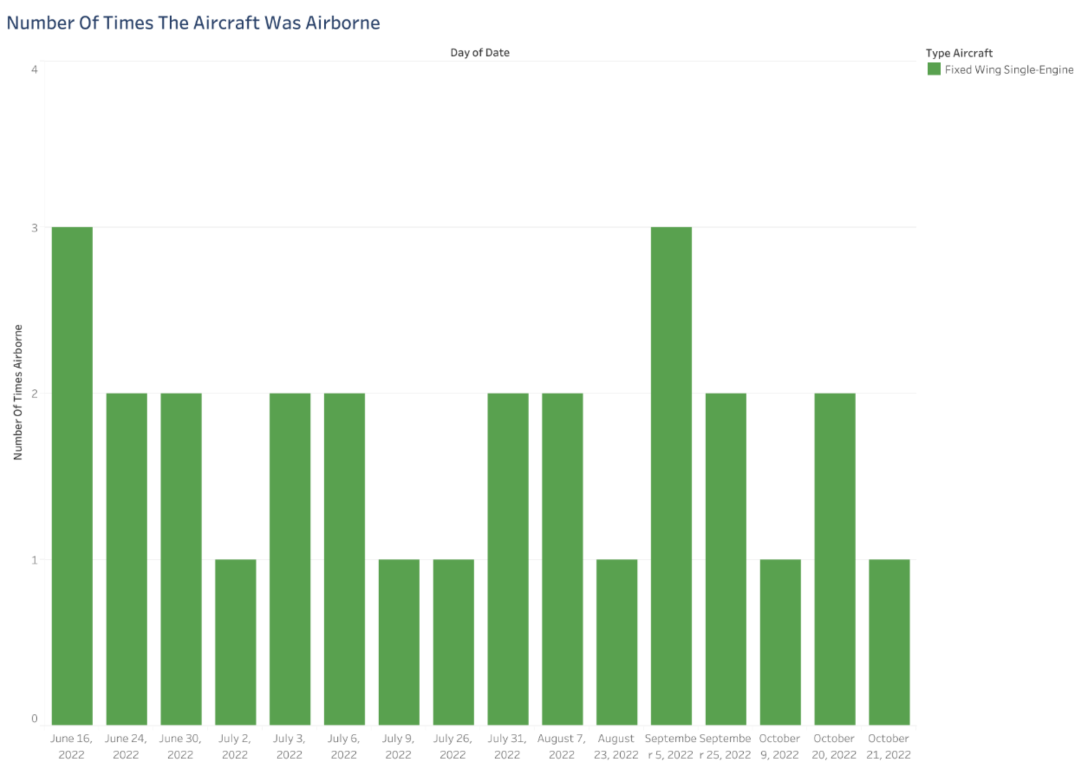
This Visualization displays the Landing gear configuration of the aircraft. Landing gears are only present for fixed engines and are categorized as S, 2S, D and 2Dwhich are acronyms for single, two single tandem, dual, and two dual tandem.

Age of Aircrafts

The next visualization shows the fly time for fixed wing single engine aircraft used in aerial firefighting with weeks of dates in 2022. According to the visualization for the given time period, the week of July 3–10 has the most flying time, clocking in at more than 220 minutes.



Flight Time for Aircrafts



Number of Times the Aircraft was Airborne

The visualization shows the frequency of airborne operations for fixed-wing, single-engine aerial firefighting aircraft on specific dates. According to the visualization, the most frequent dates the aircraft took to the air were June 16 and September 25.

Navigation Equipment Of Airports

Loc Id	ICAO Id	Runway Id	Surface Type Condition	Base Marking Condition	Edge Light Intensity	Wind Indicator
ZZV	KZZV	04/22	ASPH-G	P	HIGH	Y-L
		16/34	ASPH-G	G	MED	Y-L
ZPH	KZPH	01/19	ASPH-G	F	HIGH	Y-L
		05/23	ASPH-E	G	MED	Y-L
ZNC	-	05/23	GRAVEL	-	-	Y
ZER	KZER	04/22	TURF-G	P	-	Y-L
		11/29	ASPH-G	G	MED	Y-L
ZEF	KZEF	07/25	ASPH-G	G	MED	Y-L
Z99	-	12/30	ASPH	-	-	Y-L
Z98	-	02/20	ASPH-G	G	LOW	Y-L
Z95	-	07/25	GRVL-DIRT-P	-	-	Y
Z93	-	13/31	GRVL-DIRT-F	-	-	Y
Z92	-	04/22	TURF-G	-	-	Y
Z91	-	16/34	GRAVEL-G	-	MED	Y-L
Z90	-	15/33	TURF-G	-	-	N
Z87	-	E/W	WATER	-	-	N
		N/S	WATER	-	-	N
Z84	PACL	01/19	ASPH-G	G	MED	Y-L
Z81	-	15/33	GRAVEL-G	-	-	Y
Z78	-	E/W	WATER	-	-	N
		NE/SW	WATER	-	-	N
Z71	-	NW/SE	WATER	-	-	-
Z59	-	NE/SW	WATER	-	-	Y
Z58	-	N/S	WATER	-	-	N
Z55	-	13/31	GRAVEL-E	G	-	Y
Z52	-	09/27	GRAVEL-P	-	-	N
		18/36	GRVL-DIRT-P	-	-	N
Z48	-	15/33	TURF-DIRT-P	-	-	N
Z47	-	16/34	GRVL-DIRT-G	-	-	Y
Z43	-	NW/SE	WATER	-	-	N
Z40	-	08/26	GRAVEL-F	-	-	Y
Z33	-	E/W	WATER	-	-	N
Z25	-	11/29	TURF-GRVL-L	-	-	N
		18/36	GRVL-DIRT-L	-	-	N
Z20	-	NE/SW	WATER	-	-	N
		NW/SE	WATER	-	-	N

Navigation Equipment of Airports

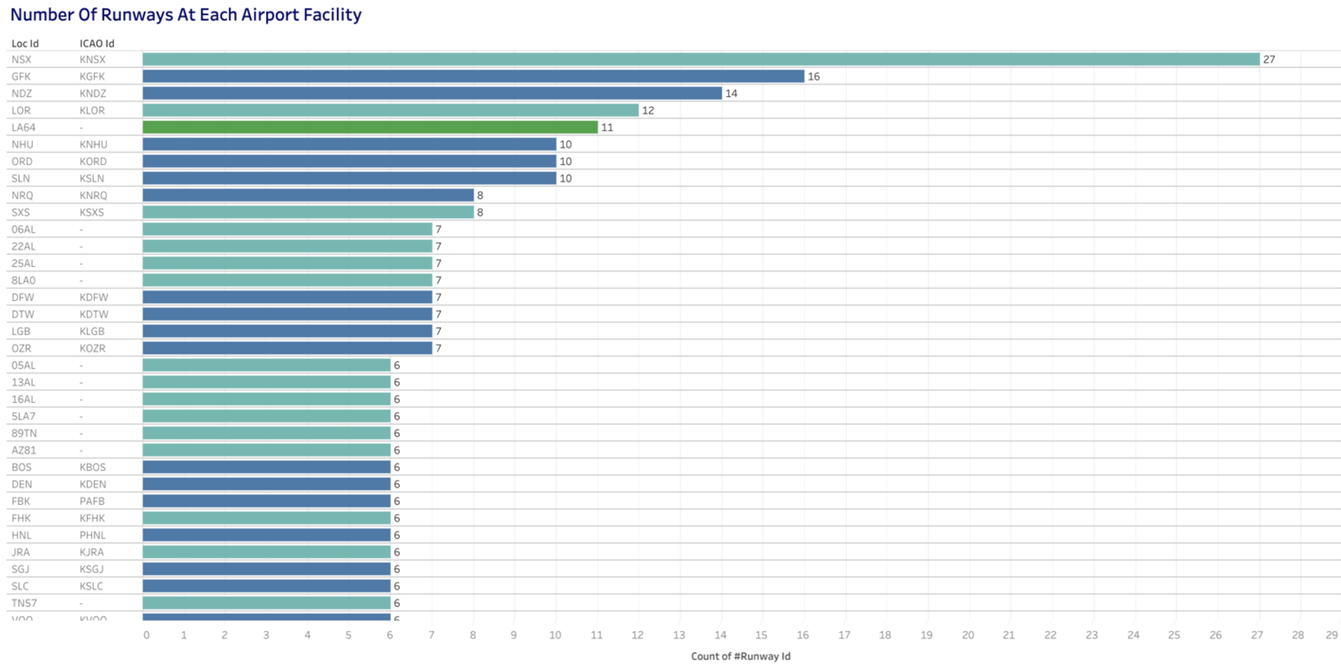
The location, runway, and surface type conditions, base marking conditions, edge light intensity, and wind indicator are all provided by the navigation equipment visualization of the Airports. This knowledge is crucial for aerial firefighting because it enables pilots to safely navigate and land in challenging circumstances.

Weight Bearing Capacity Of Each Runway

Loc Id	ICAO Id	Runway Id	WBC Single	WBC Dual	WBC Two D..	WBC Tande..
ZZV	KZZV	04/22	38	50	75	-
		16/34	38	50	75	-
ZPH	KZPH	01/19	47	69	-	-
		05/23	53	75	-	-
ZNC	-	05/23	-	-	-	-
ZER	KZER	04/22	-	-	-	-
		11/29	21	-	-	-
ZEF	KZEF	07/25	25	-	-	-
Z99	-	12/30	-	-	-	-
Z98	-	02/20	-	-	-	-
Z95	-	07/25	-	-	-	-
Z93	-	13/31	-	-	-	-
Z92	-	04/22	-	-	-	-
Z91	-	16/34	-	-	-	-
Z90	-	15/33	-	-	-	-
Z87	-	E/W	-	-	-	-
		N/S	-	-	-	-
Z84	PACL	01/19	-	-	-	-
Z81	-	15/33	-	-	-	-
Z78	-	E/W	-	-	-	-
		NE/SW	-	-	-	-
Z71	-	NW/SE	-	-	-	-
Z59	-	NE/SW	-	-	-	-
Z58	-	N/S	-	-	-	-
Z55	-	13/31	-	-	-	-
Z52	-	09/27	-	-	-	-
		18/36	-	-	-	-

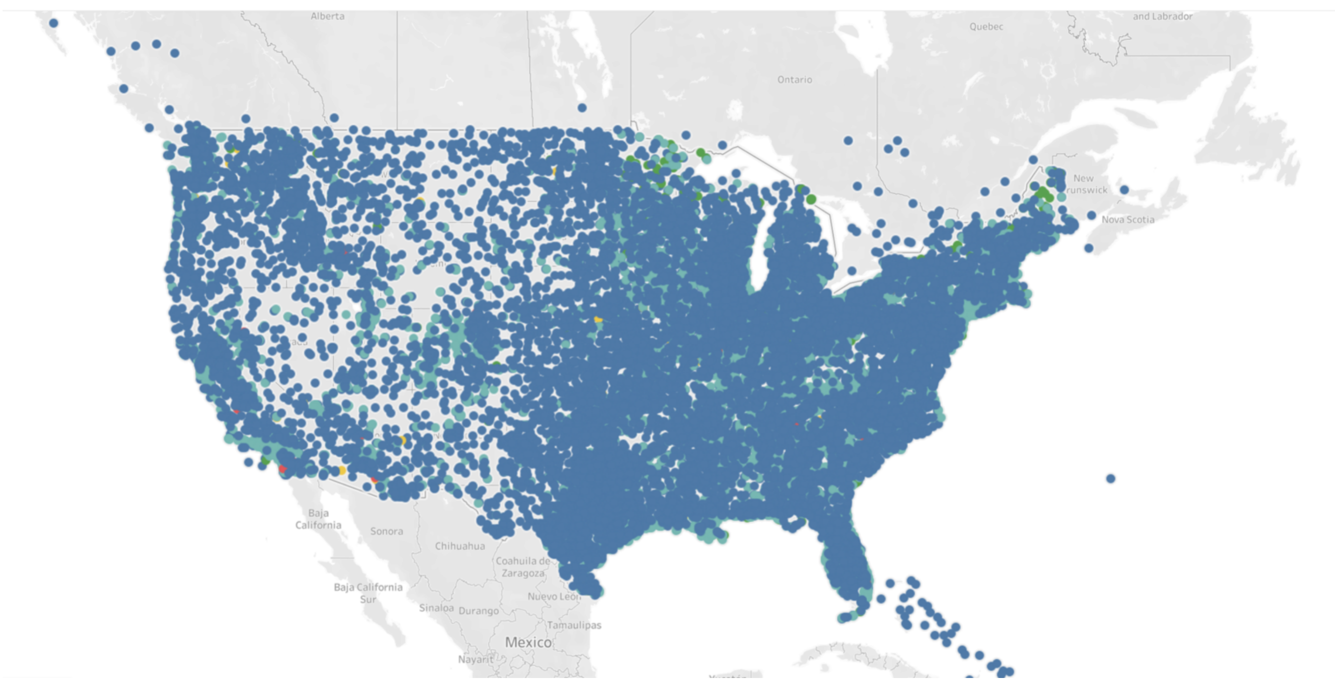
Weight Bearing Capacity of Each Runway

Along with the runway's location and runway ID, the visualization shows the runways' weight-bearing capacities. The ability to identify which runways can safely support the weight of firefighting aircraft is crucial for aerial firefighting because it allows pilots to avoid unsafe landing zones.



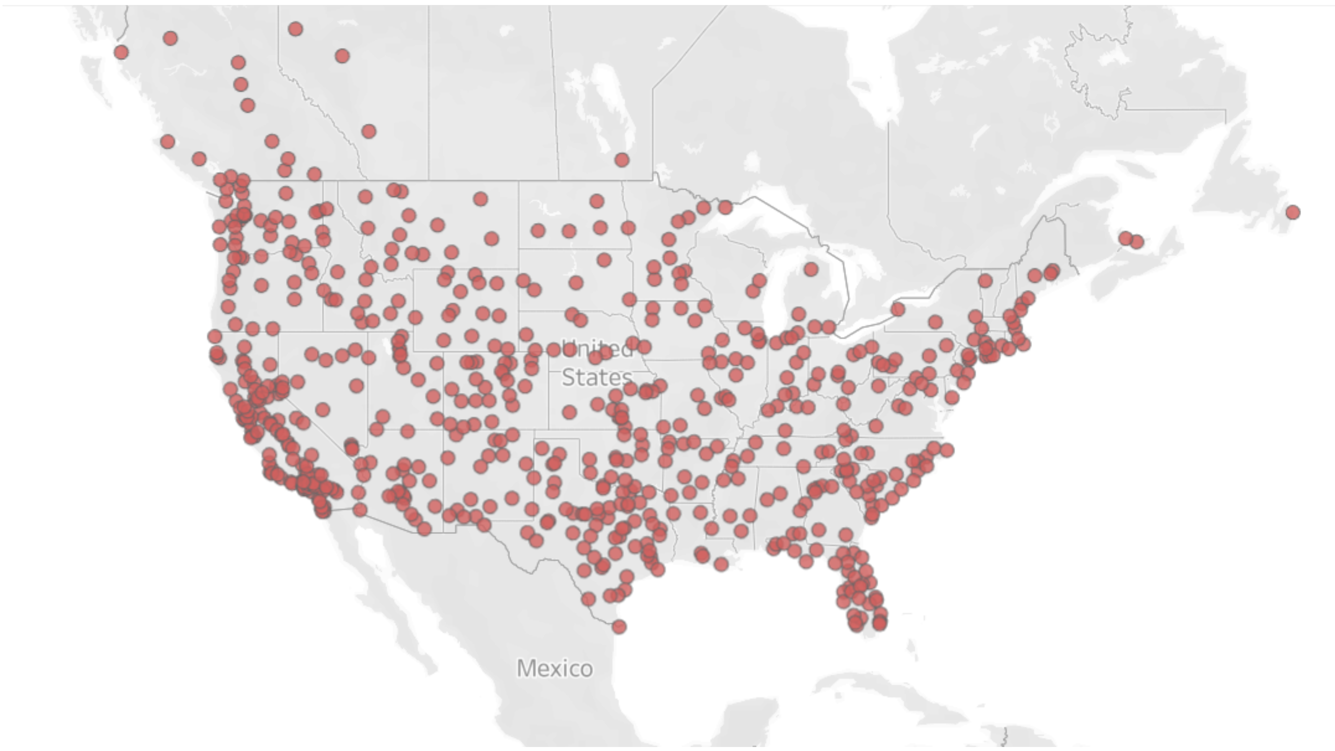
Number of Runways at Each Airport Facility

The visualization gives details about the number of runways at each airport facility. The number of runways that are available at each facility affects how many firefighting aircraft can be accommodated there, which is important to know for aerial firefighting. As a result, when planning and carrying out aerial firefighting operations, this visualization can offer firefighting teams' useful information.



Location Of Airports and Facility Type in the USA

The Location of Airports and Facility Type visualization offers details on where airports are located and the kinds of facilities offered in the USA. This knowledge is crucial for aerial firefighting because it aids pilots in finding nearby airports and facilities that can be used for logistical needs such as refueling, maintenance, and other requirements.



Fire Fighting Aircrafts Previous Landing Locations

The visualization offers details on the firefighting aircraft's previous landing locations. It gives historical data of the landing locations of the aircraft and could be used to forecasting which airports would likeli be busy in the case of fire.

Section 4: Time-Series Analysis, Forecasting, and Data Visualization

4.1 Overview

Team AgniFuego has dedicated most of our time, focus, and energy throughout Spring 4, beginning at the end of Spring 3, to becoming acquainted with Amazon Web Services' interface and learning how to use it for their large-scale project.

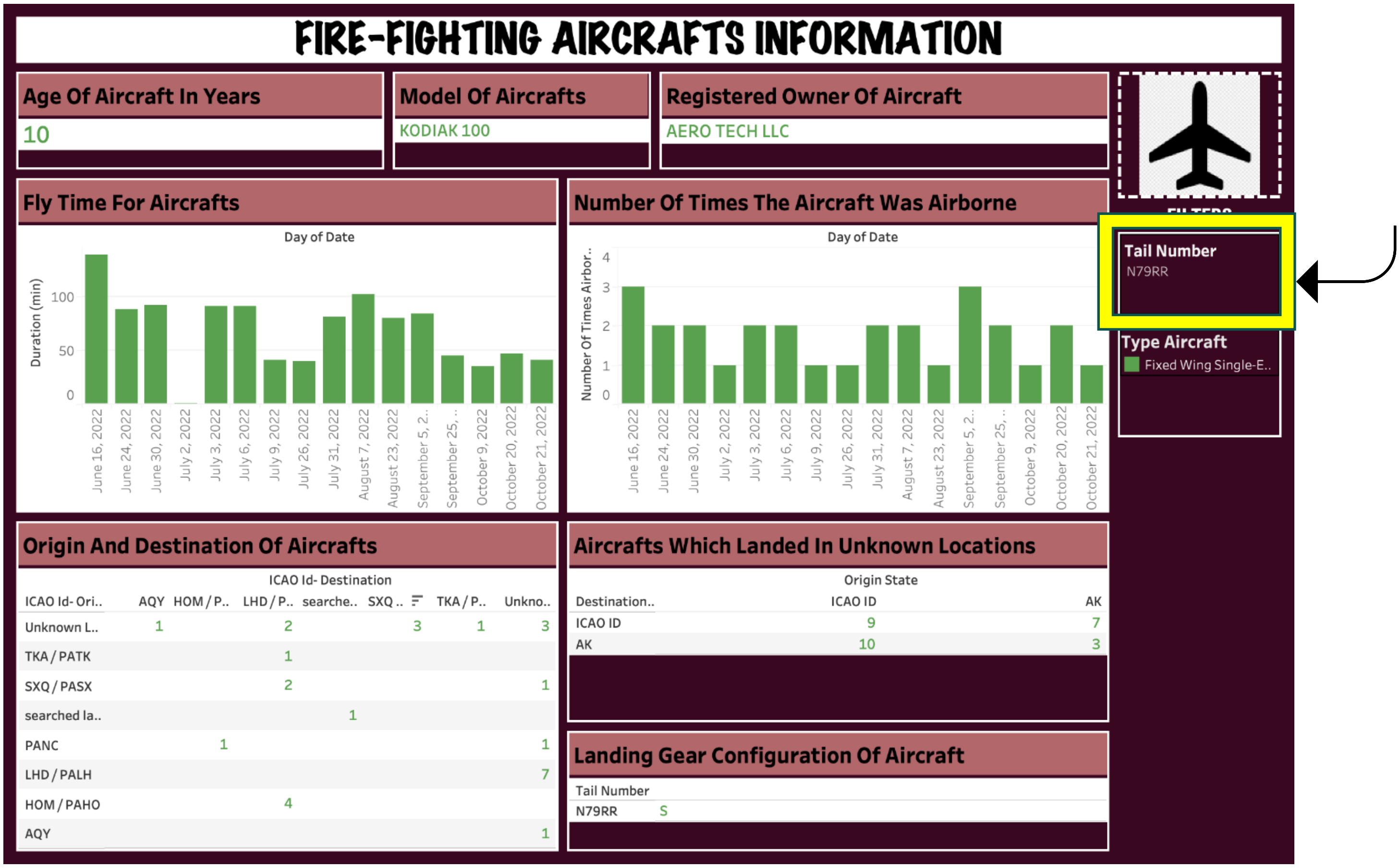
The overarching goal of this capstone project is to construct a real-time dashboard of aerial firefighting operations, capabilities, and resources across the United States. While the project does not involve machine learning algorithms or deep learning networks, they are creating numerous data visualizations to aid higher-level stakeholders and decision-makers in the United States government.

By presenting real-time information on aerial firefighting resources, the project can help prevent the spread of wildfires and minimize their impact on communities. The team's work is crucial in dealing with the public menace caused by unexpected and often uninvited infernos that spread quickly before containment, which can happen when aerial firefighting resources are mismanaged and underutilized.

Moreover, at the behest of Dr. Lance Sherry, we have also performed time-series analysis with respect to the total number of flights involved in aerial firefighting operations taken per day throughout the United States. This was done as an attempt to further assist any would be decision makers and stakeholders at the US Forest Service, FAA, BLM, CalFire, or possibly even the National Parks Service to spot patterns which are practically speaking, just not possible to detect without these analytical tools (or alternatives like them) at their disposal. Toward this end, we performed forecasts of the number of flights per day using three different time-series forecasting methods for several different days in the past in order to compare how well each of them performs in order to assess if one of them is clearly the best.

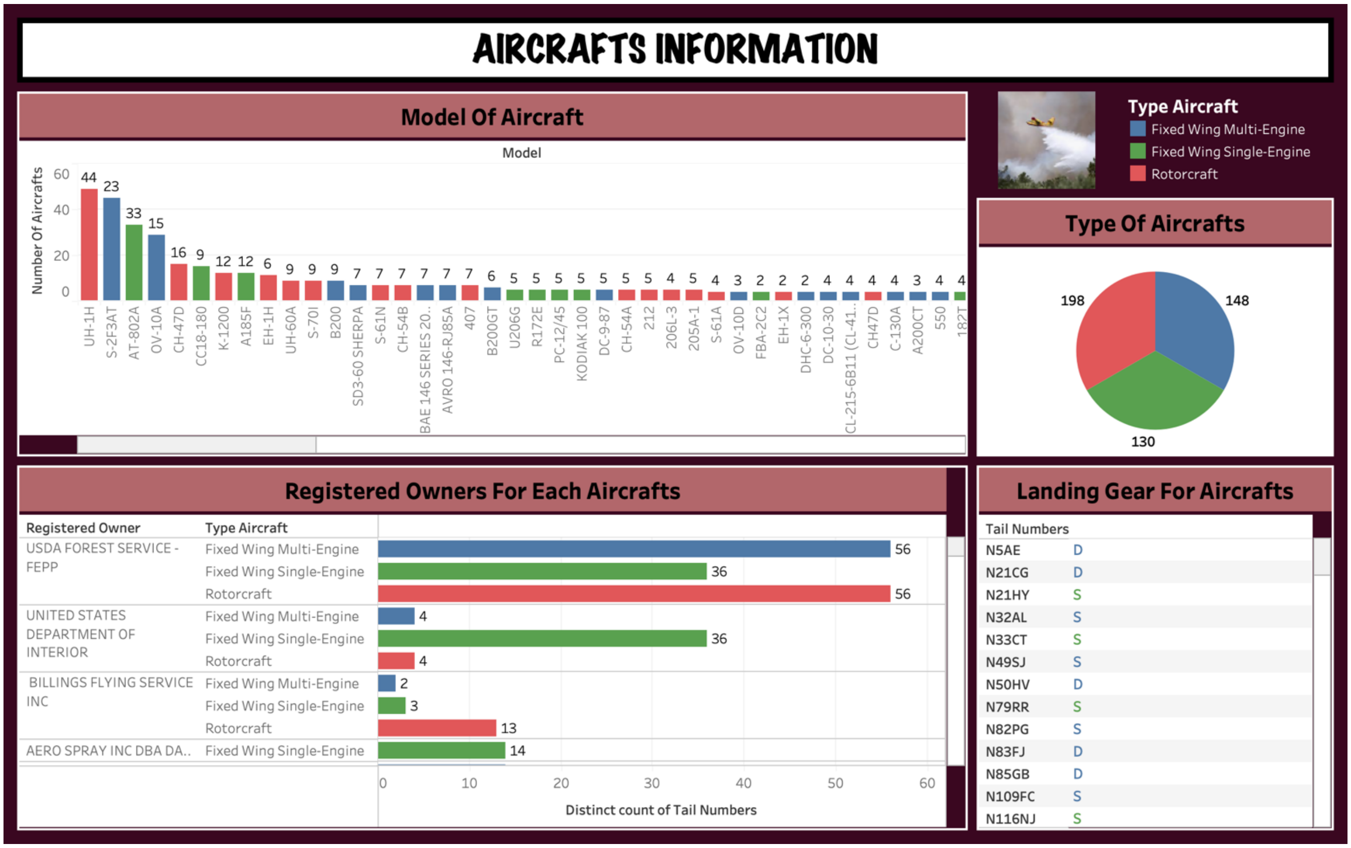
4.2 Dashboards (Assembling the Visualizations Together)

Part-1



The final dashboard is divided into 3 sheets. The filter on the right enables the user to select a particular tail number. The tail number corresponds to fire-fighting aircrafts. The first sheet the displays Aircraft Information, such as Age, Model, Registered Owner, and Landing Gear Configuration of the Aircraft, which was extracted from the FAA N-registry database. The flying time in minutes and the number of times the aircraft was airborne for a particular day are rendered into bar charts. This information was extracted from the FlightAware website. The legend color characterizes the type of aircraft, Green represents fixed-wing single-engine, Blue fixed-wing multi-engine, and Red rotorcraft. Multiple selections can be made from the filter, and the information will be displayed accordingly.

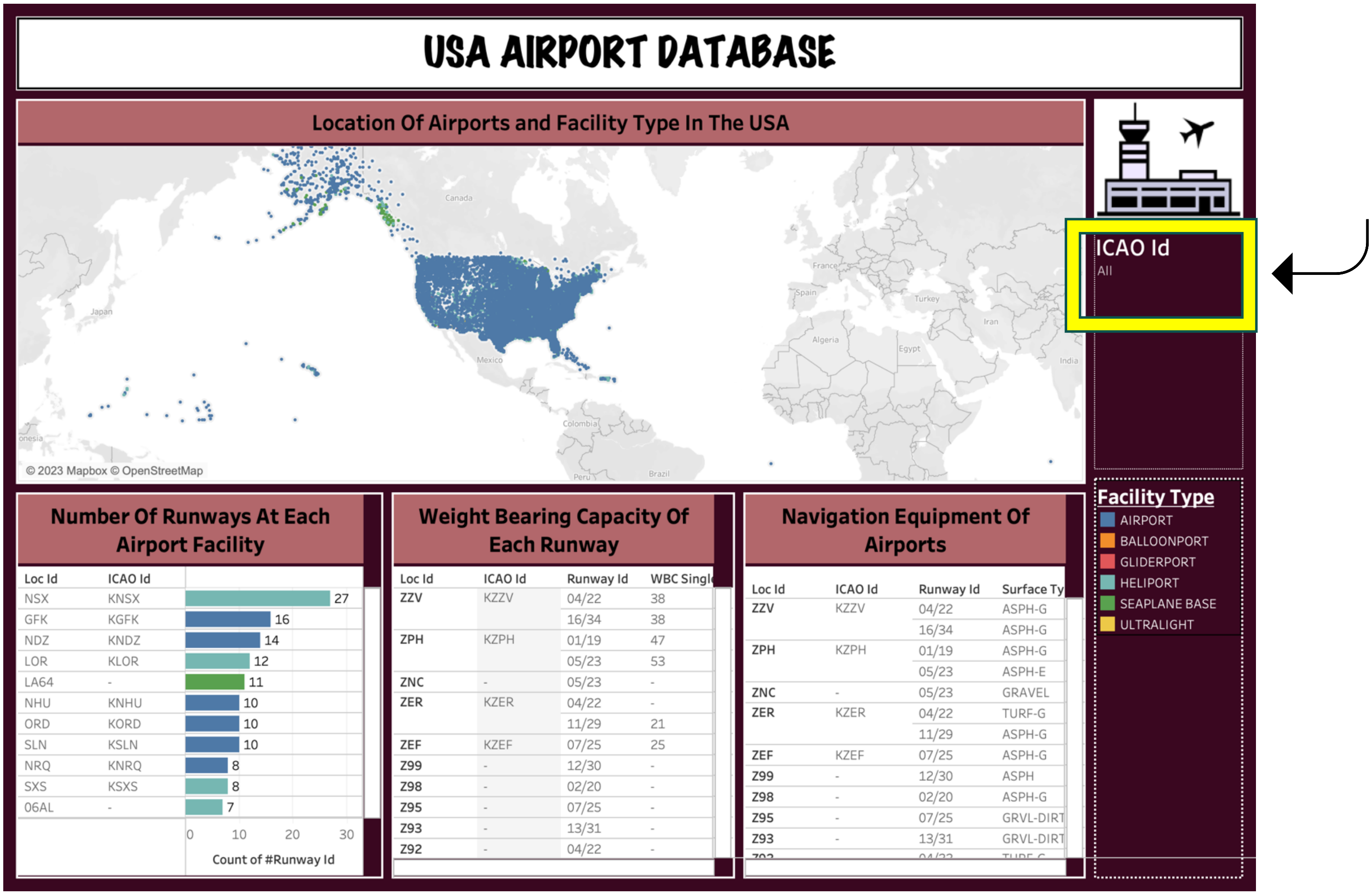
Part-2



This part of the dashboard is more descriptive rather than interactive. It summarizes many aspects of the Aircraft information. Among the fire-fighting tail numbers provided by the client when searched, most of them are rotorcrafts which are 198 in number—followed by 148 fixed-wing multi-engine and 130 fixed-wing single-engine. Forty-four of the 198 rotorcrafts are UH-IHs, sixteen CH-470s, and twelve K-1200s. For the fixed-wing single-engine, a maximum is twenty-three S-2F3ATs and fifteen OV-10As. And thirty-three fixed-wing single engines are AT-820As.

The second bar graph gives an overview of the ownership of these aircrafts. USDA Forest Service owns a maximum of 56 fixed-wing multi-engine aircraft and rotorcrafts each and 36 single-engine aircraft. USDA FEPP refers to the United Service Department of Agriculture Federal Excess Personal Property; it is a program that enables State Foresters to loan these aircrafts for the purpose of rural and aerial fire-fighting. And most of these aircraft had once belonged to the DoD (Department Of Defense). This program was first initiated in the year 1956. Since the Fixed Wing aircarfts have landing gear configurations, those have to listed to the right. Where S represents Single Landing Gear, D – Dual , 2S -Two Point Single and 2D – Two Point Dual Landing Gear.

Part-3



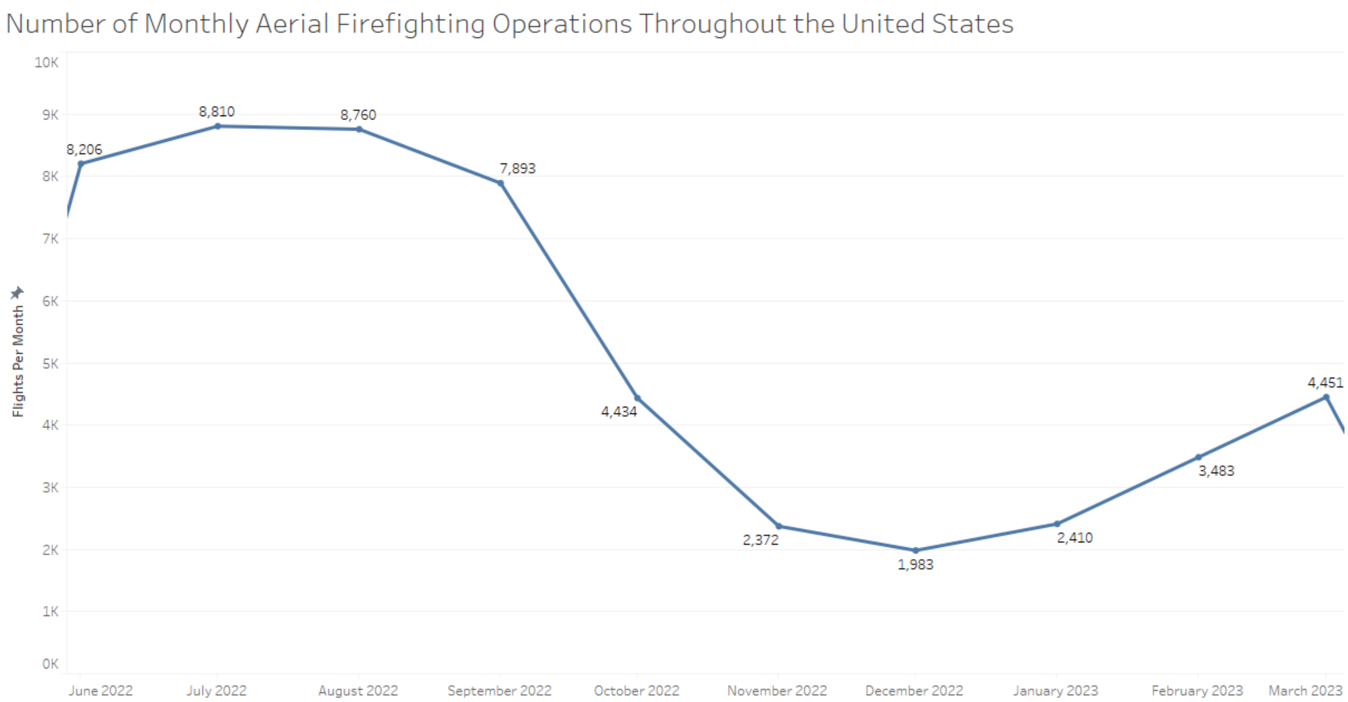
The third and final part of the dashboard is an interactive sheet that aids the user in navigating through different aspects of airport information in the United States. A selection can be made using the drop-down box on the right side of the dashboard. Multiple Ids can be selected simultaneously. The location of the aircraft is displayed on a map with the latitude and longitude information. The color of the circle describes the facility type as depicted in the legend.

Some airports have multiple runways, helipads, seaplane bases, etc. This is displayed in a bar chart, the number of runways and the type of runway for a particular facility. Since not all airports have the ICAO (International Civil Aviation Organization) Id, the Loc (Location) Id is parallelly provided with the ICAO id. The charts next to the number of runways display characteristics of these runways, such as their weight-bearing capacities and navigation equipment. The Navigation Equipment points out the surface type GVRL -Gravel ASPH -Asphalt, TURF-Grass etc. and the condition of base marking whether it is Good-G Fair-F or Poor-P. The edge light intensity is characterized by high, medium, and low. And the last column shows whether the airport has a wind indicator and if it is a lighted one. Y-L means that there is a wind indicator which is lighted. Hence this dashboard gives a complete overview of the airports in the USA.

4.3 Time-Series Analysis General Patterns, and Forecasting

Seasonality

Below is a graph showing the total amount of recorded flights per month involved in aerial firefighting operations across the United States regardless of the type of aircraft:



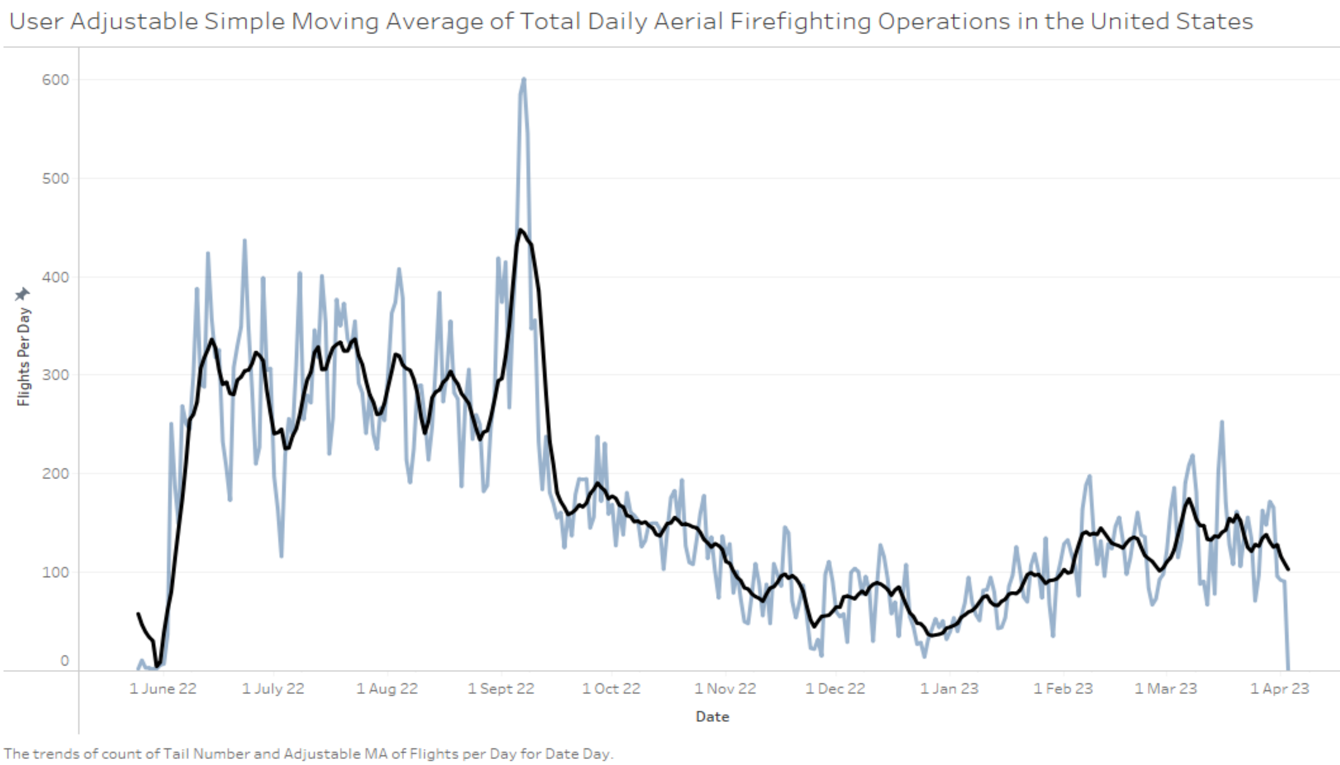
The trend of count of Tail Number for Date Month. The marks are labeled by count of Tail Number.

The graph shows that there is a noticeable increase in aircraft activity throughout the summer, with a peak in activity occurring around July and August. This is probably because there are more wildfires during the summer, and more aerial firefighting operations are needed to put out those fires. After the summer, there is a gradual decline in aircraft activity in the fall and winter, which again, is probably related to a decline in the prominence of wildfires in a similar manner.

Moving Average

The first of the three methods of time-series analysis we employ is calculating a moving average of the total number of aerial firefighting operations which take place per day. This is done in order to try to get a sense of the difference between the acute peaks which take place on the worst few days of each year and the medium-term trends in terms of how many aerial firefighting aircraft, broken down by type[2], are used to fight wildfires on a typical day, potentially empowering relevant stakeholders at the medium level (county or region wide supervisors perhaps) to make informed decisions on inventory management strategies at specific airports withing their jurisdictions. This approach could also plausibly provide a valuable lens through which to examine the intricate trade-offs involved in optimizing the number of firefighting aircraft available at any given time nationwide as well.

A chart of this moving average for all daily flights is included below:



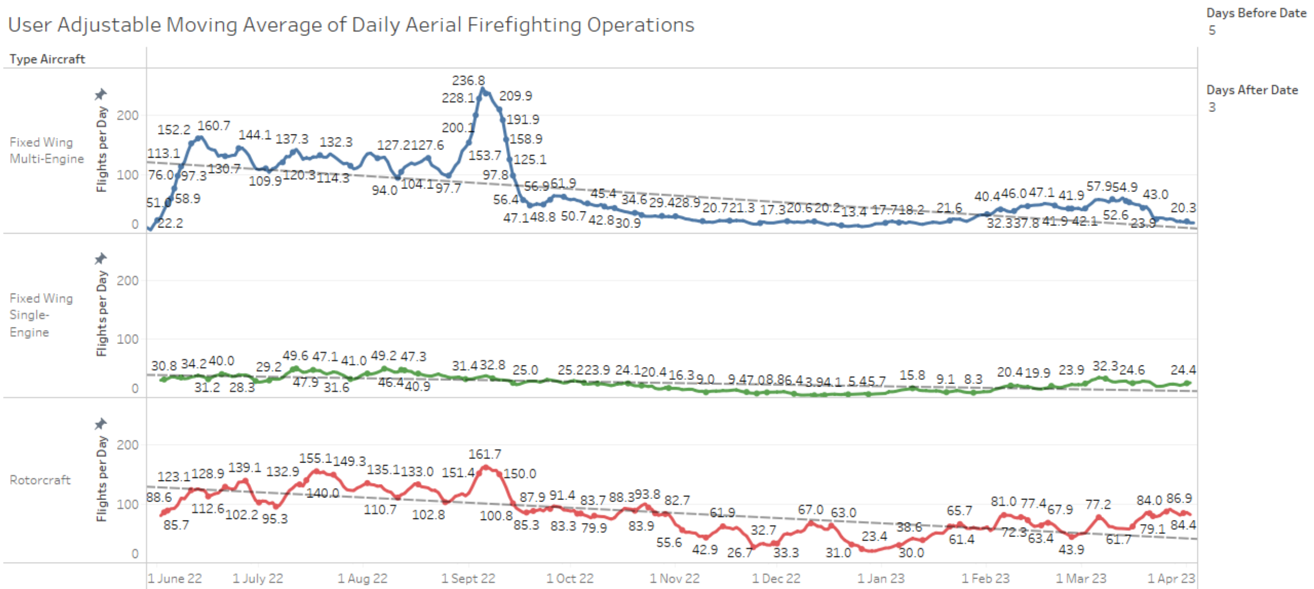
The opaque blue line is the standard line of daily flights involved in firefighting operations while the solid black line is the moving average of their daily flights. Both lines are for all aerial firefighting aircraft regardless of what type they are.

This graph is somewhat useful by itself, but in one important sense, it can be improved rather easily. The way in which this layout is non-optimal is that it is more aggregated than it needs to be. To the extent that predicting the number of overall aerial firefighting operations which are going to take place tomorrow is useful, it would be even more useful to have these predictions broken down by Aircraft Type for more than one reason. With that as the motivation behind it, we also created another line chart which is very similar to the one above, but different in two consequential ways. The first difference, it has daily flights broken down by the three broad types of firefighting aircraft the US federal government

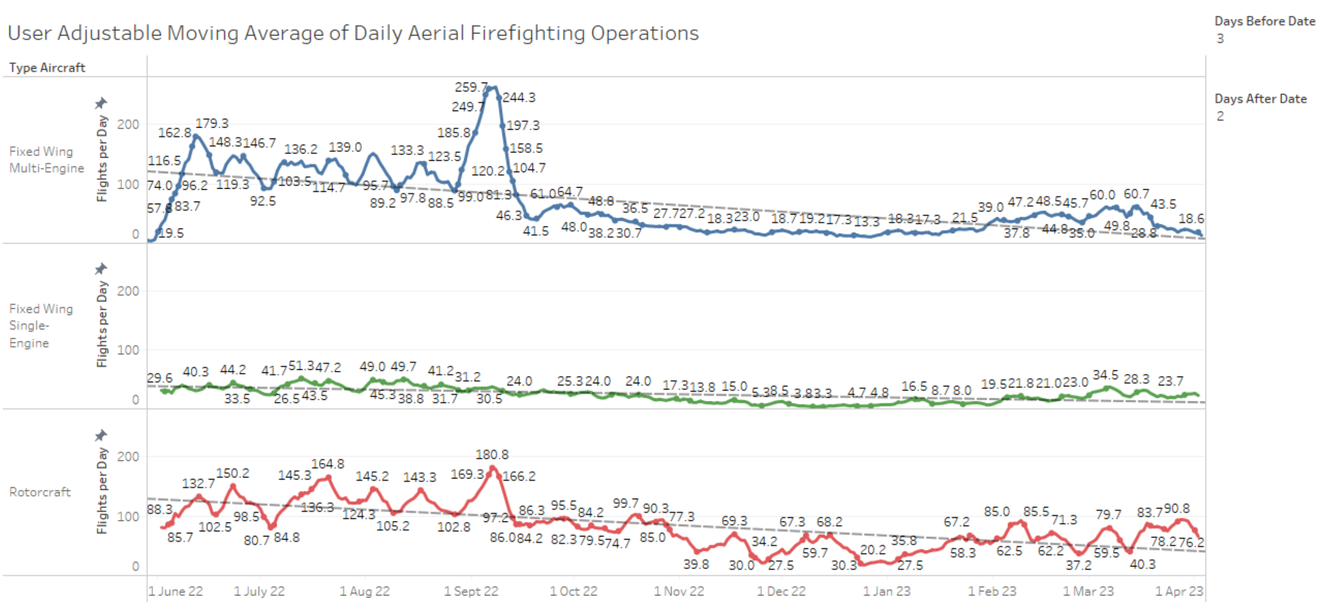
[2] The three basic types of aerial firefighting aircraft being Fixed Wing Multi-Engine, which are airplanes, Fixed Wing Single-Engine, which are also airplanes, and Rotorcraft, which are helicopters.

has at its disposal (Rotorcraft - helicopters, Fixed Wing Single Engine - propeller nosed planes, and Fixed Wing Multi Engine - proper jets or cargo planes) carrying out these firefighting operations each day; and the second difference is that each of the three line graphs partitioned by Aircraft Type only has the single line following the Moving Average of flights per day (adjustable by the user of the dashboard) included without also having the real line of raw daily flight counts behind it for comparison as is the case for the graph at the top of this page.

The graph of the Moving Average by aircraft type is included below:



This moving average uses the 5 days before and the 3 days after each date.

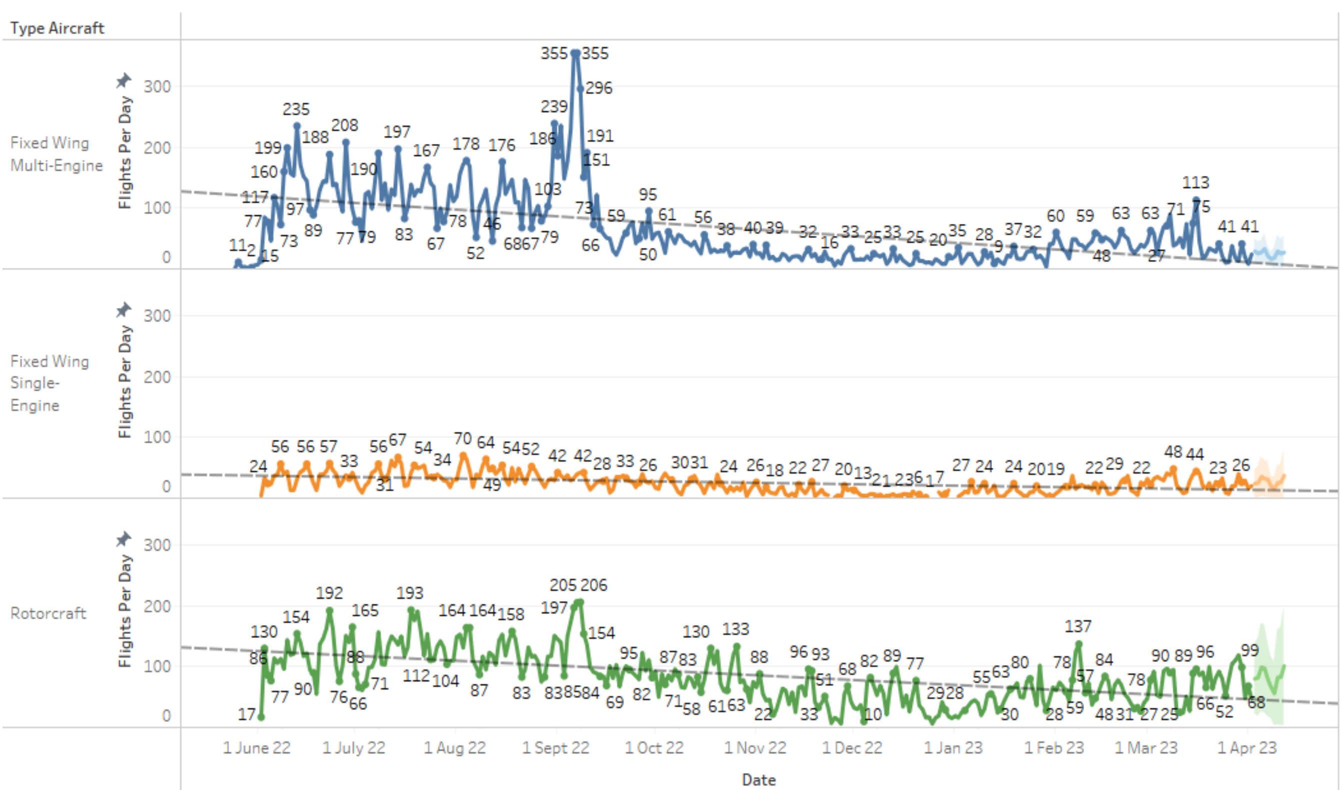


This moving average uses the 3 days before and the 2 days after each date.

Exponential Smoothing

Moving averages are useful in boosting one's ability to pick up the signal versus random noise, but unfortunately, Tableau lacks the option of using its moving average options to make forecasts with. However, by adding in linear trends, you are able to see where the mean of these MA lines are implicitly predicted as going in the future and that it a suitable workaround given that Tableau does provide the option of Exponential Smoothing forecasts as well. Exponential Smoothing models iteratively forecast future values of a regular time series of values from weighted averages of past values of the series ([How Forecasting Works in Tableau - Tableau]()). The reason this smoothing is said to be exponential is because the weight assigned to the degree of influence of values further and further back in time goes down exponentially. Now arises the issue of seasonality because wildfires are extremely seasonal in the United States. Fortunately, the Exponential Smoothing forecasts Tableau provides do allow for the inclusion of seasonal effects and our forecasts included multiplicative seasonal effects in them, they are depicted in the graph below:

Flights Per Day from May 2022 to April 2023 with a 1 Week ES Forecast, and Linear Trendlines



The trend of count of Tail Number (actual & forecast) for Date Day broken down by Type Aircraft. Color shows details about Type Aircraft and Forecast indicator. The marks are labeled by count of Tail Number. The view is filtered on Type Aircraft, which keeps Fixed Wing Multi-Engine, Fixed Wing Single-Engine and Rotorcraft.

SARIMA

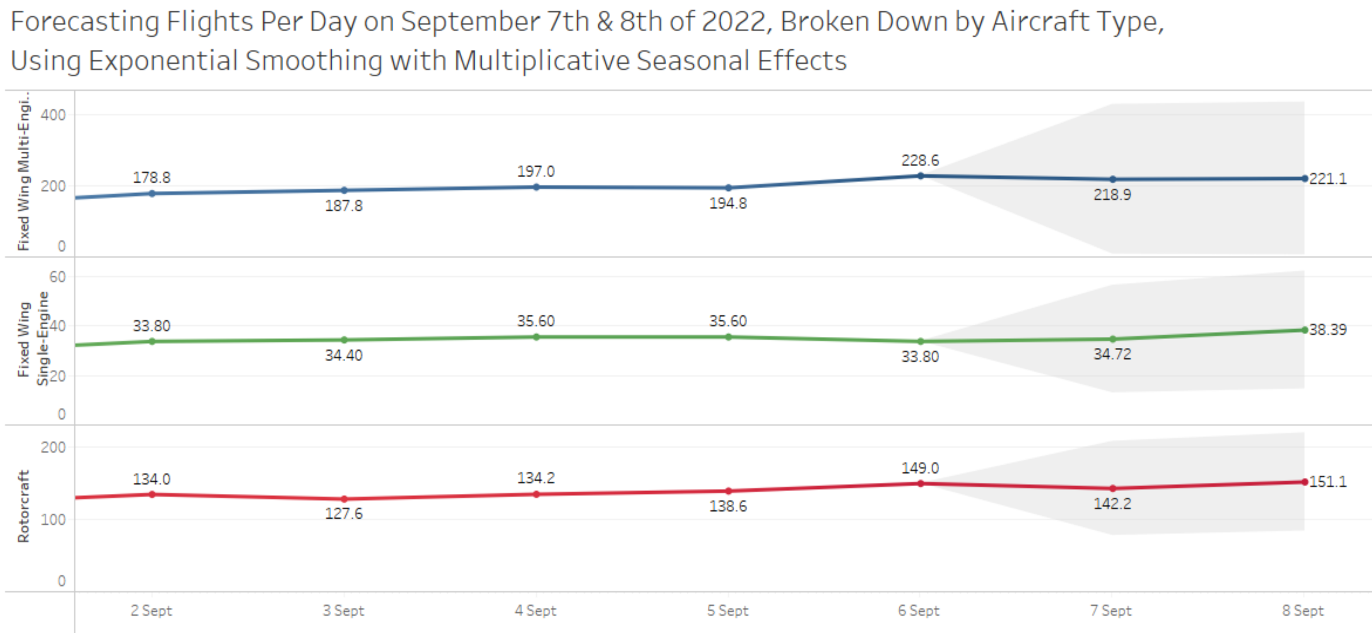
While Exponential Smoothing is a solid method of forecasting for many different purposes in the context of time-series datasets, it is not a panacea. As previously mentioned, for our context, a moving average predictive method would be preferrable here, but what specific type? For predicting the total number of daily flights, a Seasonal Autoregressive Integrated Moving Average (abbreviated as (S)ARIMA) forecast would likely make the best predictions according to the econometrics and predictive analytics literature. The (S)ARIMA model's strength lies in its ability to capture both linear and seasonal dependencies in time-series data, making it particularly well-suited for addressing the complexities of aerial firefighting operations. By employing the (S)ARIMA model, in principle, we should be better able to understand the historical patterns of wildfire activity and anticipate future trends, ultimately enabling more effective distribution of resources and enhanced preparedness. But being best in principle is not enough, so we compare the results of these three forecasting methods in the next sub section.

4.4 Comparing the Performances of MA vs ES vs SARIMA on Specific Dates

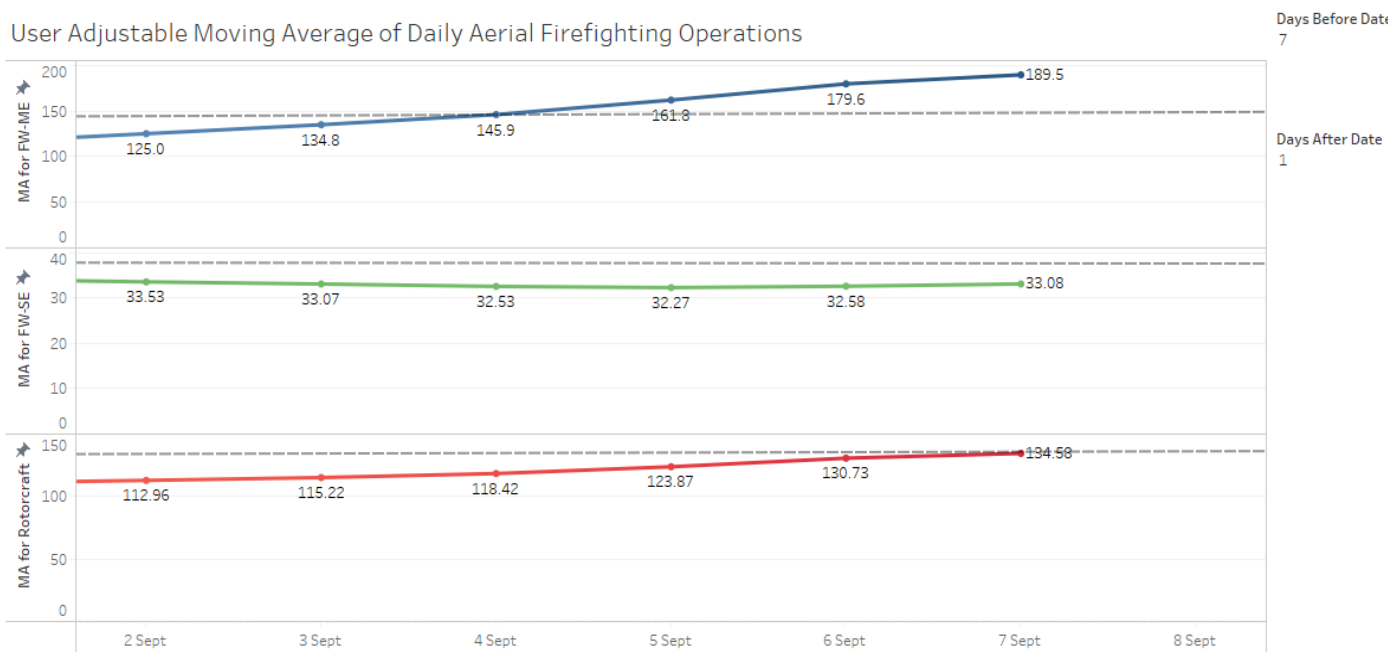
September 7th & 8th of 2022:

This pair of dates were chosen because they were the 2nd and 3rd days of Fairview Fire in California which ignited on September 5th and was not deemed as fully contained until September 16th (Fairview Fire Now 94% Contained; All Evacuation Orders and Warnings Lifted, 2022). When comparing which of several candidate forecasting methods would be the best choice to use for a complicated real-world application and that application is of such consequence that getting it right is most essential is when the local and global maxima of the phenomenon occur, it is crucial to explicitly include a head-to-head comparison of each candidate forecasting model on those dates, but in a way that gives each of them a fair chance at getting close. Think of this as a stress test.

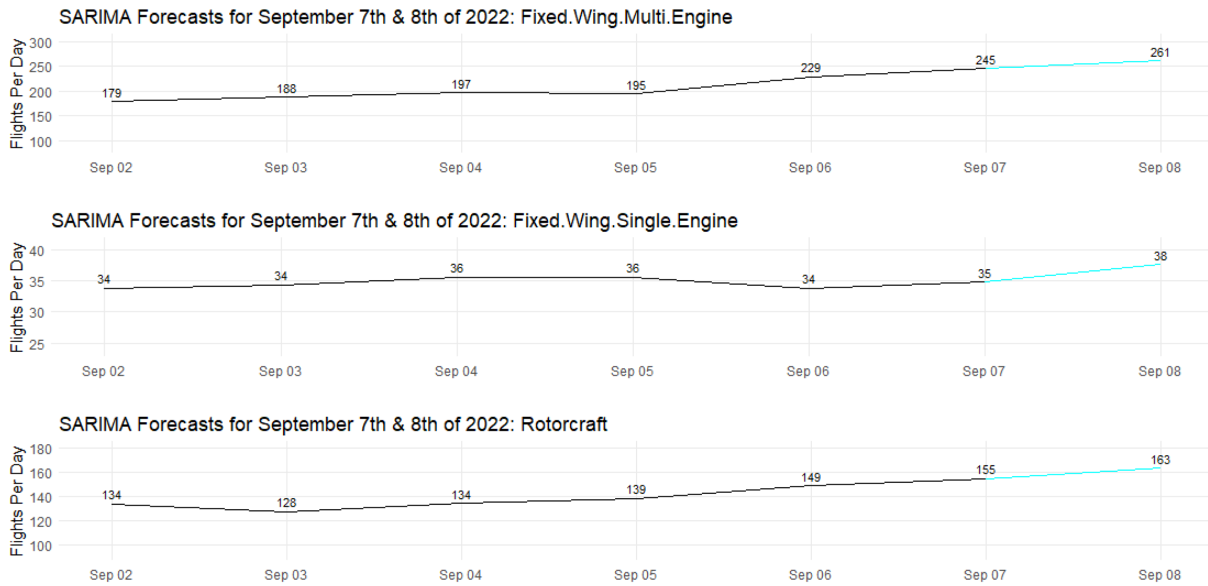
The first graph is the forecast made using an Exponential Smoothing function with multiplicative seasonal effects using all daily observations in our dataset which starts on May 25th of 2022. But, for the sake of enabling direct comparisons with the graphs for MA and SARIMA below, only the 5 dates before September 7th & 8th are included along the horizontal axis:



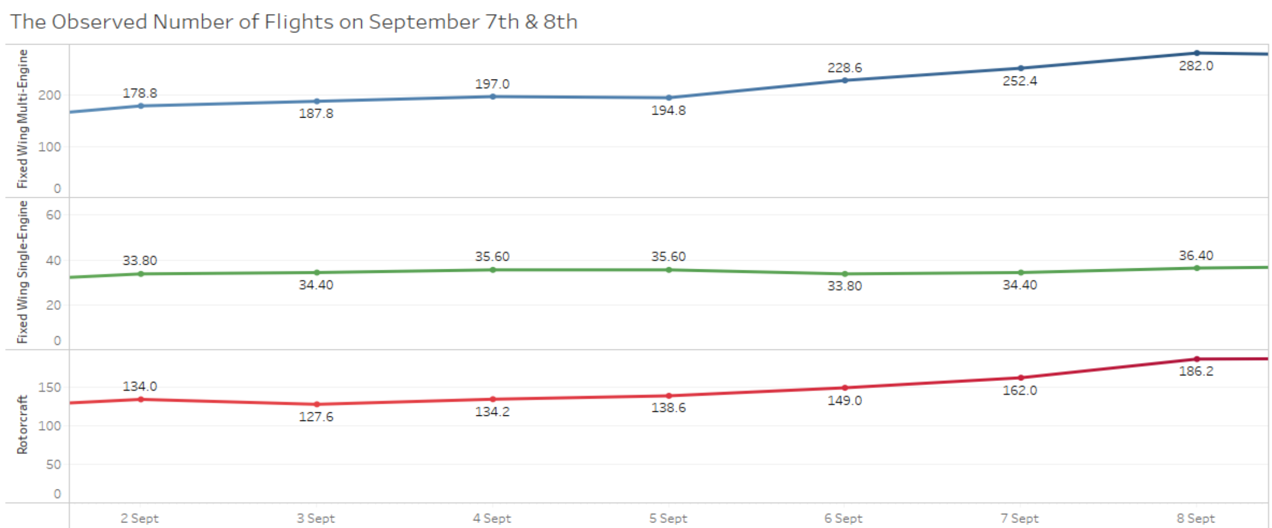
The second graph is the implicit forecast made by following the linear trendline of the 7-day moving average of flights per day (broken down by aircraft type as always) where the 7-day window is the 7 days before each date which is chosen because this is often seen as a reasonable calculation window. However, the number of days before and after each date to smooth the lines by is up to the user to decide in our Time-Series Analysis of United States Aerial Firefighting Operations Dashboard which can be clearly seen by the Days Before Date and Days After Date user adjustable parameters in the chart below:



Comparable line charts for each SARIMA for that same time period:

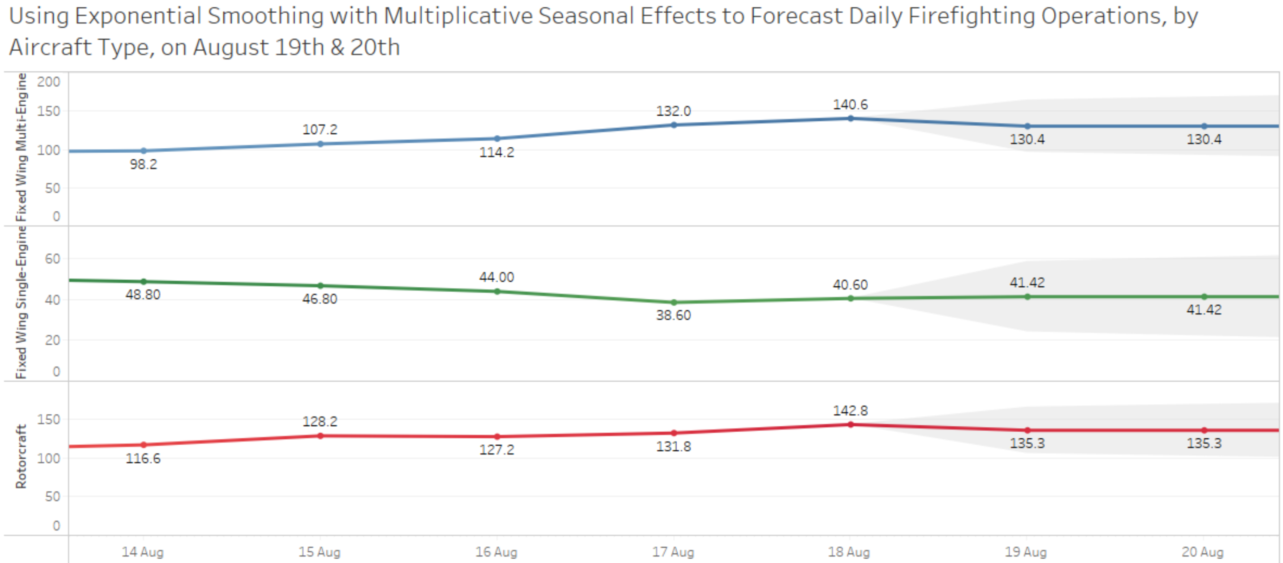


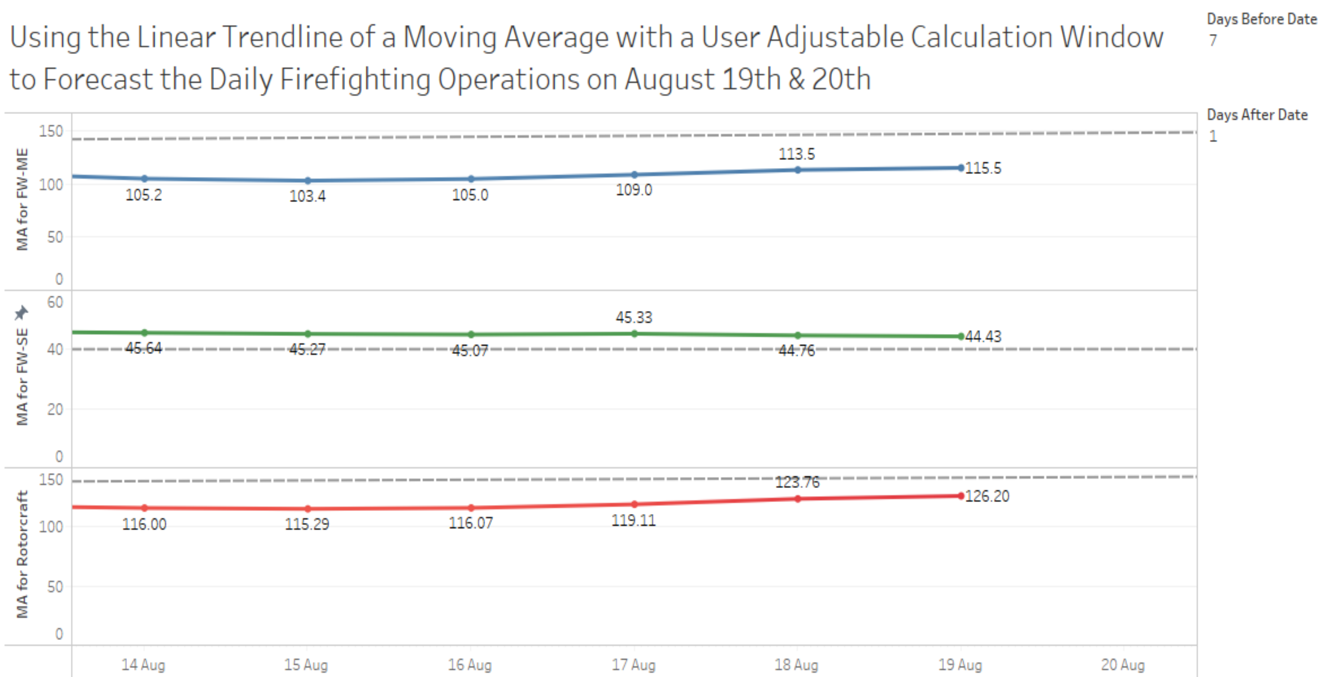
Here is the true number of daily flights from September 2nd through September 8th, 2022:

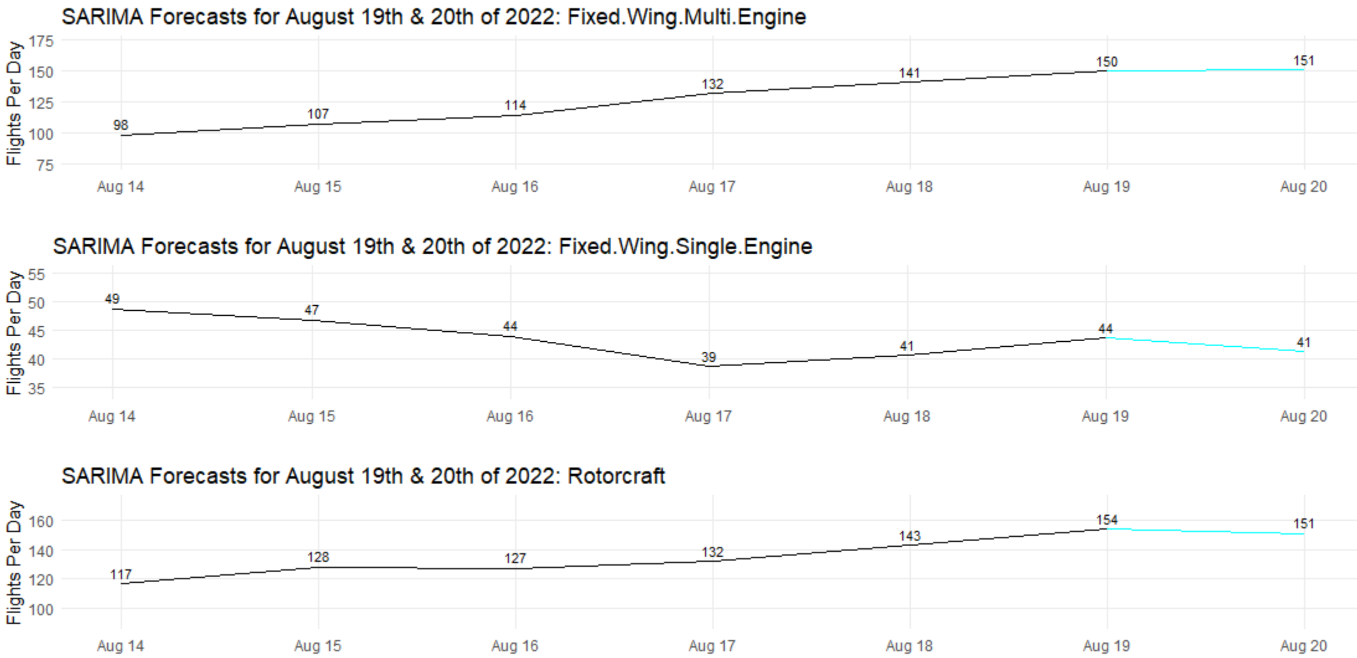


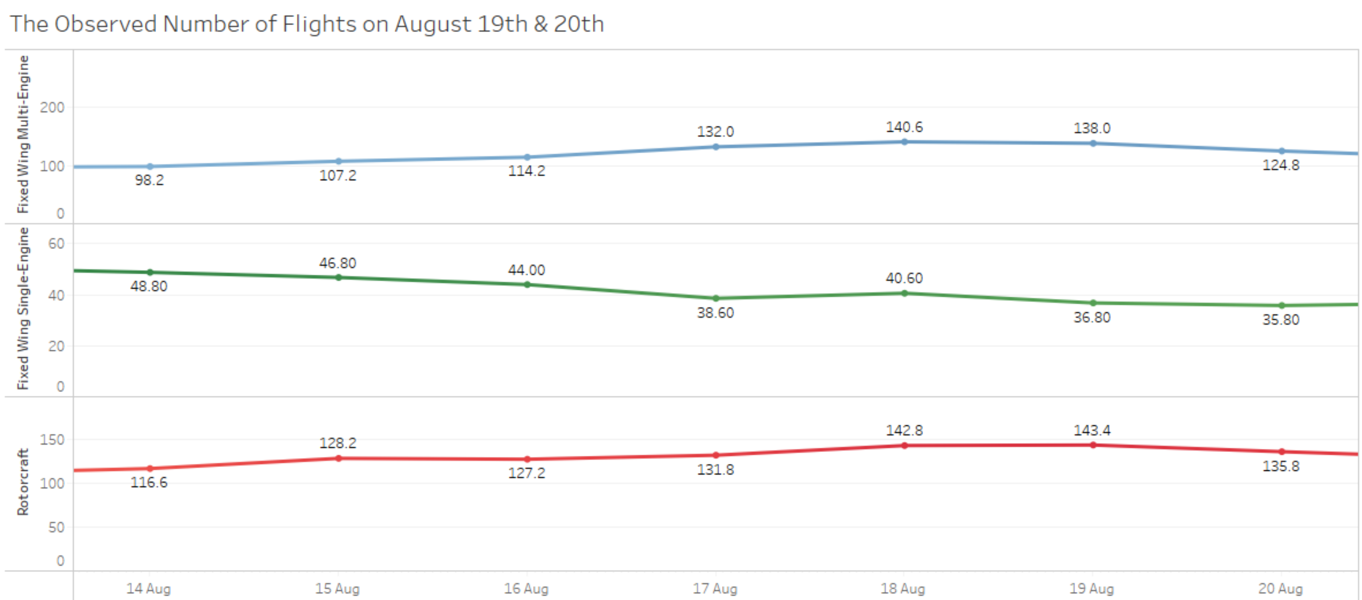
Winner for September 7th & 8th: SARIMA

The same four charts for August 19th & 20th:









Winner for August 19th & 20th: Exponential Smoothing with Multiplicative Seasonal Effects

Forecasts for a handful of other discrete pairs of consecutive two-day periods in the past were also made and the comparison of those forecasts with their actual observed values are included in Appendix E.

Which forecasting method performed the best for each pair of dates both here and in Appendix E was determined in Microsoft Excel using spreadsheet calculations and the resulting table is included below:

Date	Actual			SARIMA			Exponential Smoothing			Moving Average		
Type	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**
25/6/22	150	45	153	-8%	-3%	5%	-7%	4%	-25%	13%	-40%	-19%
26/6/22	144	41	139	-8%	-8%	9%	-27%	18%	-21%	23%	-37%	-8%
Mean % Errors by Forecasting Model & Aircraft Type				8%	5%	7%	17%	11%	23%	18%	39%	13%
				7%			17%			23%		
3/7/22	86	26	104	35%	-3%	7%	31%	28%	19%	101%	20%	32%
4/7/22	85	23	90	44%	-15%	20%	25%	41%	37%	112%	28%	59%
Mean % Errors by Forecasting Model & Aircraft Type				40%	9%	13%	28%	34%	28%	106%	24%	45%
				21%			30%			59%		
16/7/22	135	52	139	-1%	5%	-2%	1%	-7%	3%	19%	-38%	-9%
17/7/22	129	43	136	-3%	8%	-9%	-16%	4%	-2%	27%	-29%	-5%
Mean % Errors by Forecasting Model & Aircraft Type				2%	6%	5%	8%	5%	3%	23%	33%	7%
				5%			5%			21%		
19/8/22	138	37	143	9%	20%	7%	-6%	11%	-6%	3%	11%	-2%
20/8/22	125	36	136	21%	15%	11%	4%	15%	-1%	15%	15%	6%
Mean % Errors by Forecasting Model & Aircraft Type				15%	17%	9%	5%	13%	3%	9%	13%	4%
				14%			7%			9%		
7/9/22	252	34	162	-3%	3%	-4%	-13%	3%	-12%	-40%	15%	-17%
8/9/22	282	36	186	-7%	4%	-12%	-22%	4%	-19%	-46%	7%	-27%
Mean % Errors by Forecasting Model & Aircraft Type				5%	4%	8%	17%	4%	16%	43%	11%	22%
				6%			12%			25%		
11/1/23	16	14	38	14%	32%	-5%	8%	4%	11%	1%	-65%	13%
12/1/23	18	15	40	10%	32%	-8%	-1%	12%	13%	-23%	-74%	5%
Mean % Errors by Forecasting Model & Aircraft Type				12%	32%	6%	4%	8%	12%	12%	69%	9%
				17%			8%			30%		
1/3/23	36	19	36	-10%	-26%	-1%	7%	-4%	13%	-75%	-79%	10%
2/3/23	43	21	46	-23%	-29%	-17%	-7%	-6%	-8%	-81%	-81%	-15%
Mean % Errors by Forecasting Model & Aircraft Type				17%	27%	9%	7%	5%	10%	78%	80%	12%

	18%	7%	57%
	12%	**12%**	**32%**

A different table that has all of the actual values for all those dates and also the point estimates forecasted on those dates by each forecasting method rather than their percentage errors is also included in Appendix E.

Section 5: Findings

This project was both challenging and tremendously satisfying for our Team. The FAA partner provided data with multiple options and explanations. Our team extracted aircraft and airport data from multiple websites such as FAA, FlightAware, and many more, then analyzed and made recommendations on where to go based on what the data was telling us. The fact that no previous research used the data was both interesting and frightening.

The most significant challenges involved in scraping, cleaning, sorting/organizing our flight operations data, and generating our analysis were mostly achieved via Python. These tasks took up half of the time we spent working on this project. That is not necessarily a negative; however, the Team believes we are now successfully presiding over a significantly cleaner and more usable dataset than anything we were able to find publicly available that we could now send back to the FAA ready for further use any further research projects similar in subject and scope to ours.

The analysis we performed generated a few of what we believe will new insights for the FAA and the US Forest Service, which had already been lurking in their data beforehand.

We extensively analyzed aircraft types, airport locations, and flight tracking information.
Finally, the Team was able to contribute useful insight to our FAA partner, and we believe that our work will be the first step for designing even more complex dashboards that will help in the preparedness for aerial fire fighting.

In terms of patterns and trends in daily flight operations, we found that there are, as one would expect, strong seasonal trends in terms of the need for aerial firefighting throughout the year with the need going up greatly during the summer months and down every winter.

Section 6: Summary

The final dashboard gathers the necessary information from three data sources (FAA Nregistry, FAA airport database, FlightAware) and helps provide insights for US Forest Service.

The first part of the dashboard gives information about the age, model and ownership of the aircraft. The aircraft shockingly range from 1 year old to 81-year-old. There are four aircraft are above 70 years old, and all are fixed-wing single-engine. Among all the aircraft there are 43 UH-1H rotorcrafts, 23 S-2F3AT multi-engine, and AT-802A single-engine, And the maximum number of aircraft 56 fixed-wing multi-engine, 56 rotorcrafts, and 36 fixed-wing single-engine are owned by USFDA.

The second part of the dashboard focuses on the flight operations information, the routes a particular aircraft is taking, how many times was it airborne on a particular day, and also the fly time in minutes. When analyzing the activity of the aircraft separately based on its type (fixed wing single-engine, fixed-wing multi-engine, rotorcraft), the activity for fixed-wing multi-engine and rotorcrafts show unusual spikes in the time series plot.. When studied further, these spikes translate to dates on which the wildfires occurred in 2022. For fixed-wing single-engine there is a cyclic pattern. These aircraft operate at a bare minimum on weekends. Safe to say it would be quite a nuisance if a wildfire occurs on a Sunday.

The next part of the dashboard focuses on airports. Subdivided into 4 parts, the first part gives the geographical coordinates of the airport facility. The number of runways and type (helipad, seaport, glider port, runways etc.) And then the information on weight bearing capacity for each runway, and finally the navigation equipment present at each facility.

And finally, during the implementation of three different time series methods (Exponential Smoothing, Moving Averages and SARIMA) to forecast the daily flight operations, the Exponential Smoothing and SARIMA models both made good forecasts which were far superior to those made by simple moving averages and equal to each other overall. However, SARIMA performed best by far when it counts, namely, during a crisis, when there is a serious wildfire ablaze and on that basis, it is to be preferred. Both the typical, medium and longer terms averages and the maximum observations within the dataset for the number of Fixed Wing Multi-Engine airplanes and Rotorcraft (Helicopters) engaging in aerial firefighting operations on a daily basis are much higher than the number of Fixed Wing Single-Engine airplanes on any given day.

On the whole, the dashboard puts together various jig-saws of aerial firefighting from various sources together into a single frame that will help an extensive group of people, such as Authorities in individual state fire control to Randy Moore, Chief of Forest Service, strategically tackle wildfires.

Section 7: Future Work

Our findings suggest several potential follow-up tasks or project next steps. The data can first be further analyzed to find patterns and trends, particularly in relation to aircraft types, weather, and location. This can aid in the creation of predictive models that can aid in the scheduling of resources and the making of decisions during firefighting operations.

Second, it's important to investigate additional data sources that can supplement the information the FAA provides. To obtain a more complete dataset, this may entail working with other agencies and organizations engaged in firefighting operations.

The potential for using machine learning and other advanced analytics techniques to analyze the data must also be investigated. By doing so, it may be possible to uncover patterns and insights that conventional data analysis techniques might not make immediately obvious. But also utilize what we learned in terms of forecasting flights per day, namely that SARIMA and Exponential Smoothing are both better than Moving Averages, but SARIMA forecasts are better during acute out of control wildfires.

Appendix

Appendix A: Glossary

Term	Definition
Landing Gear (S,D)	S-Single D-Dual
Base Mark Condition(G,F,P)	G-Good , F-Fair, P-Poor
Surface Type Condition	CONC-Concrete, ASPH-Asphalt or Bituminous Concrete, TURF- Grass, Dirt-Soil, E-Excellent,G-Good, F-Fair, P-Poor, L-Failed Conditions
Edge Light Intensity	High-H, Med-Medium, NTSD-Nonstandard lighting system, None-no,L-low
ICAO ID	International Civil Aviation Organization Id
Wind Indicator	Y-Yes, N-No, L-Illuminated
Urban Sprawl	The expansion of urban areas into formerly rural areas
Fire Suppressants and Retardants	Chemicals used to extinguish or slow the spread of wildfires
Data Integration	The process of combining data from different sources and formats into a unified data set
Incident Commanders	Individuals responsible for overseeing the overall response to a wildfire incident
Streaming Data	Data that is generated continuously and delivered in real-time
Moving Average	This method of forecasting calculates the average of the counts over the past 7 days and uses that value as the forecast for the next day. As each new count becomes available, the oldest count in the 7-day window is dropped and replaced with the new count.

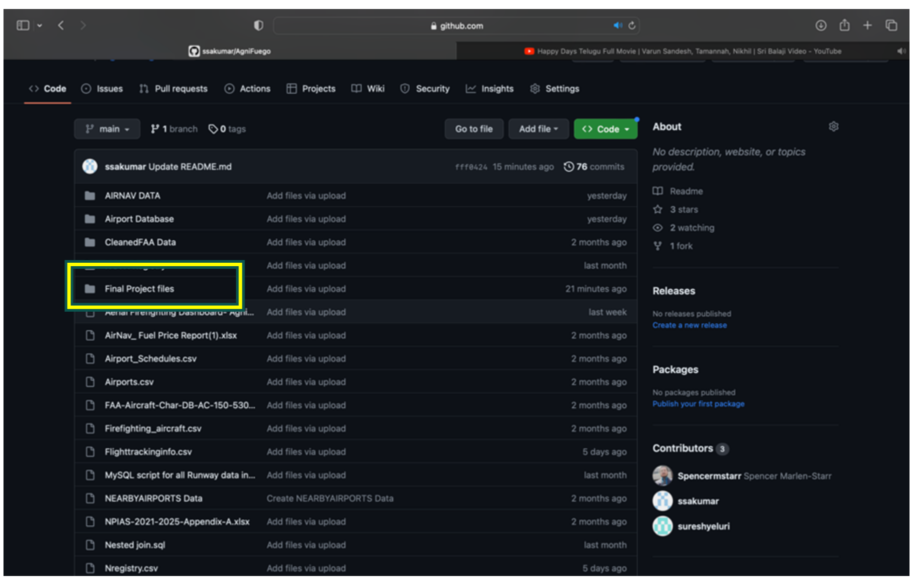
Exponential Smoothing	This method of forecasting assigns exponentially decreasing weights to each count in the past 7 days, with the most recent counts given the highest weights. The weighted sum of the past 7 counts is then used as the forecast for the next day.
ARIMA	A more complex forecasting method that models a time series as a combination of autoregressive (AR) and moving average (MA) processes.
SARIMA	This is an extension of the ARIMA model that takes into account seasonal patterns in the data. It is particularly useful when the counts exhibit a recurring pattern over time.

Appendix B: GitHub Repository

Overview

The project emphasizes building a real-time dashboard that helps tackle wildfires across the USA. To achieve this goal, three data sets must be extracted. First is the FlightAware data, which provides information on the flight paths for each day starting from May 2022 until today. It also gives the duration of the flight to reach its destination. Next is the information on the aircraft. This would define the registration details, type of aircraft, etc., from https://registry.faa.gov/aircraftinquiry/Search/NNumberInquiry. And last would-be information on the airports and runways. This is just a downloadable Excel file from the link below. https://adip.faa.gov/agis/public/#/airportSearch/advanced. Initially, we did not discover this file, so many attempts were made to scrap the data. The GitHub repository consists of many versions of Python codes, which were scripted to gather the datasets. The Final Project Files folder consists of the functional Python codes that extract the required data and the datasets.

GitHub Repository Link: https://github.com/ssakumar/AgniFuego



Link to Dashboard:

https://public.tableau.com/views/FinalDashboardForAerialFireFightingPreparedness-Agnifuego/Fire-FightingAircraftsInformation?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Appendix C: Risks

Sprint 1 Risks

Risk Name	Description	Probability	Impact	Mitigation
Scope Of the Project	The project deals with aerial-fire-fighting of the northern hemisphere and is a vast topic and scope of it is huge	high	high	Limiting the scope to just North America
Budget	Since most of the data is protected by FlightAware there is a subscription fee to access the information	medium	medium	With help of Prof. Lance Sherry, we were able to do it free of cost
Web Scraping	Since this is relatively new to all the team members.	low	low	We must train ourselves in web scraping
Cost	Cost of using AWS services	Medium-high	high	Leveraging the AWS resources is crucial for effective data cleaning, hosting and presenting an optimal real-time dashboard

Learning AWS	Since most of us are not familiar with AWS	Medium	High	Self –study and mentoring from professor if much need

The team has identified four potential risks for sprint 1: project scope, budget, web scraping, and learning AWS. The team identified the project's risks correctly and has offered workable mitigation strategies.

Given the vast subject of aerial firefighting in the northern hemisphere, the project's scope has been identified as a high-risk factor. To lessen the impact of this risk, the team has suggested limiting the project's scope to just North America. The team will be able to produce better results if they can narrow their focus and define the project's scope.

Due to the cost of gaining access to data from FlightAware, the budget has been classified as a medium-risk factor. However, the team has been able to address this risk by making use of the tools provided by Professor Lance Sherry to gain access to data for free, which is a practical way to lessen the financial impact of this risk.

Web scraping risk has been deemed a low-risk factor. The team has suggested learning web scraping, though, to lessen the impact of this risk. This strategy is reasonable because it will help team members gain more knowledge and skills while lowering the likelihood that web scraping-related problems will arise.

Finally, given that most of the team members are unfamiliar with AWS, learning AWS has been identified as a medium-risk factor. The team has suggested reducing this risk through independent study and, if necessary, faculty mentoring. This strategy makes sense because it will enable the team to acquire the knowledge and abilities required to utilize AWS effectively while minimizing the impact of this risk.

Sprint 2 Risks

Identifying the datasets: While the code for extracting data from flightaware.com was provided by our client Dr. Lance Sherry. The other two datasets had to be extracted via python code. While extraction of N Registry was a smooth sail. The data on runways was difficult to extract as the layout of the pages varied and it was rather difficult to obtain. Moreover, multiple records were present for a single Location Id. And using APIs was not accessible as the permissions were restricted. Further risks are listed below.

Risk Name	Description	Probability	Impact	Mitigation
Multiple Records	"MMPR" has 4 airports with the same id.	High	High	Have to figure out a way.
API	Extraction Of Airport Runway Information	High	High	Study and train in APIs
Unstructured Data	All Data on Airport Runways automatically gets shoved into a single cell next to the airport ID when scraped and downloaded and a csv file	High	High	Lots of hours and effort spent in Excel recreating the proper structure for this data
Missing Records	Some flight operations are not kept track of	Low	Medium	Cross Referencing data on the same or similar things from varying sources

We had wasted a lot of time figuring out onto how to extract runways information and finally clean it. But ultimately, we were able to find a downloadable excel file in Faa.gov. If we had focused on finding the data rather than extracting in python, we would have had more time working on cleaning of the datasets.

s

Sprint 3 Risks

Risk Name	Description	Probability	Impact	Mitigation
[meaningful]	[brief]	[high, medium, low]	[high, medium, low]	[brief]
Accessing the S3 bucket and granting permissions is difficult.	We're having trouble configuring and granting access for our S3 bucket to be publicly accessible.	High	High	As soon as possible, resolve the problem.
Python runtime errors when deployed in via an AWS Lambda function	When deploying a Python script to AWS Lambda, runtime errors occur.	High	Medium	Switch to using an EC2 instance instead
Tableau Public Visualizations Moving to AWS Dashboard	To use existing visualizations, connect a personal Tableau account to an S3 bucket using an AWS Athena instance.	Medium	High	Using AWS Quick sight, replicate similar visualizations.
Getting full MySQL functionality for querying our S3 Bucket	Some SQL commands are too complex for Athena to handle, preventing the creation of the Runways data table.	Medium	High	Use AWS RDS in conjunction with Athena to enable more extended MySQL use.

While attempting to access Flight Aware Data, an HTTP 500 error occurred.	Flight aware may restrict frequently accessed IP addresses, generating an HTTP 500 error.	Medium	Medium	Restart the EC2 instance to receive a new public IP address.

- Difficulty Accessing S3 Bucket and Granting Permissions: This issue must be resolved as soon as possible so that the S3 bucket may be used efficiently. One method is to double-check the permissions to confirm they are valid before testing the bucket for accessibility.
- Python Runtime Errors in AWS Lambda Function: If you get runtime errors when deploying a Python script in AWS Lambda, you might consider using an EC2 instance instead. This will give you more control over the environment, making troubleshooting any issues that may develop easier.
- Tableau Public Visualizations Were Moving to AWS Dashboard: If you're experiencing trouble connecting a personal Tableau account to an S3 bucket using an AWS Athena instance to use existing visualizations, you might want to try utilizing AWS quick sight to replicate similar visuals.
- Adding complete MySQL capabilities to our S3 bucket: If some SQL statements are too complex for Athena to process, preventing the Runways data table from being created, you can use AWS RDS in conjunction with Athena to enable more extended MySQL usage.
- If Flight Aware blocks frequently accessed IP addresses, resulting in an HTTP 500 error, restart the EC2 instance to acquire a new public IP address, allowing you to access the data without trouble.

Sprint 4 Risks

Risk Name	Description	Probability	Impact	Mitigation
[meaningful]	[brief]	[high, medium, low]	[high, medium, low]	[brief]
Moving Averages and Forecasting Models created in Tableau	Since seasonal changes could not be added the forecasting models are not accurate	High	High	Moving to R to forecast Models accurately
Not being able to create the proper Time-Series graphs in Tableau	There is a chance that Tableau Desktop cannot create charts showing SARIMA forecasts	High	High	Create SARIMA forecast charts in R
None of the three forecasting method being very accurate	SARIMA, MA, & ES all perform poorly	Medium	Medium	Report this finding and suggest the need for further work
Not being able to publish a Tableau Workbook with Analytical Expressions extension	It is possible that I will not be able to publish a Tableau Workbook with an extension that allows it to run R code to Tableau Public	Medium	High	Just use ggplot2 to create the graphs in RStudio instead, and include them in the report & ppt, as well as the scripts used to create them in our GitHub Repo.

Sprint 5 Risks

Since this is our last and shortest sprint, we are done with the development and analysis parts of the project. The only risk faced was that the Analytics Extension in Tableau only works locally, it can't be published online. So, we decided to add the graphs manually to the dashboard.

Appendix D: Agile Development

Scrum Methodology

We found the Scrum methodology to be comparatively simple to adopt because it offered a structured method of project management and permitted more frequent communication and collaboration among us. We held scrum meetings five days a week, which helped to keep the team members updated on the project's development, issues encountered, and tasks that needed to be finished. This method made sure that everyone on the team was on the same page and assisted in spotting any potential obstacles early on.

The YouTrack tool was crucial to the efficient management of our project. We were able to keep track of tasks and issues more effectively. YouTrack also served as a good, central location for all project-related data and improved team communication. The tool was useful and simple to use, and it assisted in keeping the project on schedule.

The importance of having a clear understanding of the project requirements at the outset is one of the lessons we learned from this project. This is because it helps the team members establish a clear and shared vision to ensure that everyone is working toward the same goal and meets the requirements of each sprint.

Overall, we as a team found the scrum methodology very effective in managing our aerial firefighting project, and YouTrack was an asset in successfully managing our project.

Sprint 1 Analysis

The team concentrated on data scraping from five different data sources during Sprint 1 to support the project's ultimate objective of developing a dashboard to aid in decision-making for efficient aerial wildfire firefighting. In order to comprehend the needs and demands of the client, the team identified the user stories.

The sources included FLIGHTAWARE, AIRNAV DATABASE, NPIAS DATABASE, FAA DATABASE N-Registry, and FAA NAVIGATION AIDES DATABASE. The team was able to successfully extract the necessary data, including flight numbers, aircraft types, locations of departure and arrival, information about the airport, engine specifications, and more. Since some data sources included complex formats, the task of scraping the data proved difficult. To overcome these challenges and avoid reaching their limits, the team divided the scraping into smaller parts.

Sprint 2 Analysis

This Sprint was the longest and most intense of the Sprints, as this is where the majority of data extraction took place. In this Sprint, the team focused on extracting the data using the selenium soup package in Python and preparing the data for analysis. Although visualizations were not scheduled to start until Sprint 4, the Team started working on the initial exploratory analysis to understand the data. The task to identify and collect additional data sources from merging with the primary data source continued into this Sprint as an understanding of the data continued to evolve. The creation of calculated fields such as Aircraft Age, calculated by subtracting the current year 2023 from the manufacturing year, was initiated in this Sprint. Drafts of the project report and presentation continued throughout the Sprint.

Sprint 3 Analysis

During this Sprint, the team had to put in a lot of effort and time to complete a challenging project. The main focus of this project was to create real-time dashboards using Tableau and work with Amazon Web Services (AWS) to collect and analyze data related to aerial firefighting operations and resources. The team used MySQL, Python, and Tableau tools to preprocess, analyze, and visualize the data. Drafts of the project report and presentation continued throughout the Sprint.

The team had to use AWS services like Athena, Glue, and RDS to complete the ETL process for airport and runway datasets. The team worked tirelessly to collect, clean, and process the data. Finally, they were able to create dashboards that showed real-time data related to firefighting aircraft in the United States.

Overall, the team successfully completed the project and gained valuable insights from the data they analyzed. The project was challenging, and it required the team to work collaboratively using various tools and approaches to achieve their goals.

Sprint 4 Analysis

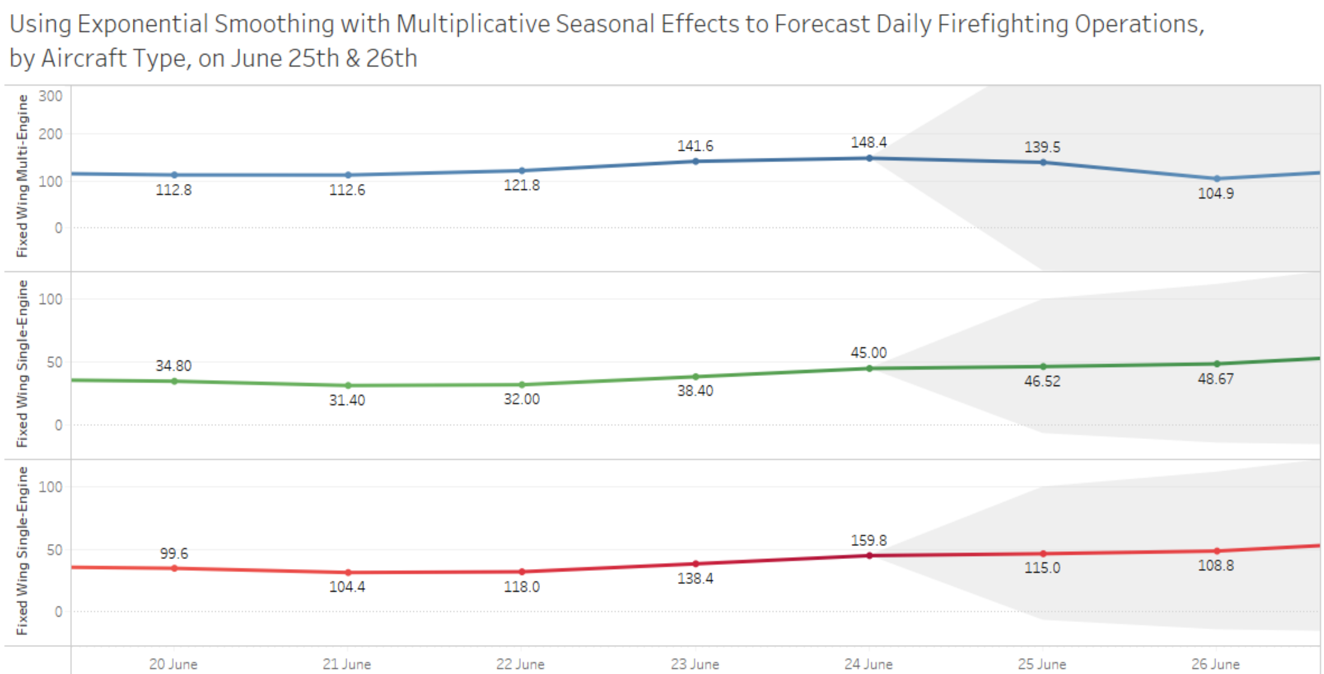
In terms of the analytical work done and progress, we made during Sprint 4, it was all about Forecasting our Time-Series data on daily aerial firefighting operations nationwide broken down by aircraft type. Our three types of aircraft here were Fixed Wing Multi-Engine (airplanes with more than one engine), Fixed Wing Single-Engine (airplanes with only one engine), and Rotorcraft (helicopters). Almost all of the analytical work done in this sprint was done by Spencer. He used Tableau, Excel, and R to complete the analysis and forecasting and the main lesson learned while doing so was that if another team is to follow up on this project, it would be better to do all of this time-series analysis and forecasting in R or Python and not really try to do any of it in Tableau because it seems easier to do in Tableau at first, but it ends up being much more tedious and less easily replicable in a reasonable amount of time.

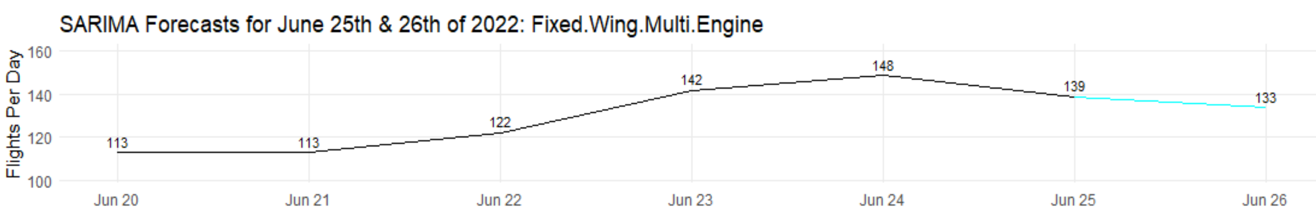
Sprint 5 Analysis

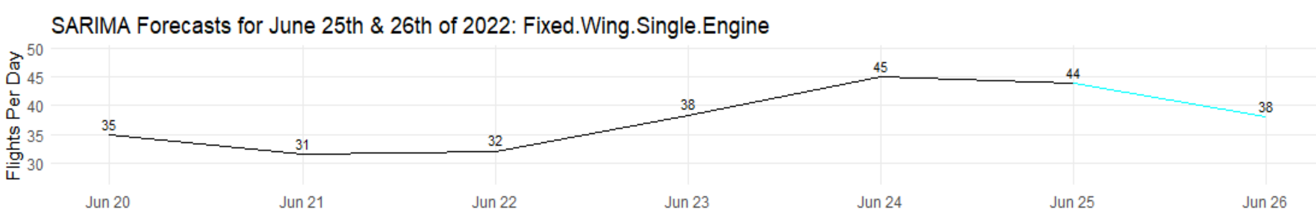
During the last sprint, our team made good progress in implementing key agile development practices such as sprint planning, daily stand-up meetings, and sprint review and retrospective. We focused on multiple tasks including time series anomaly detection, completing the report, making necessary changes and updates to the dashboard, and preparing for the final presentation. We are on track with our project timeline and met all the deadlines as planned.

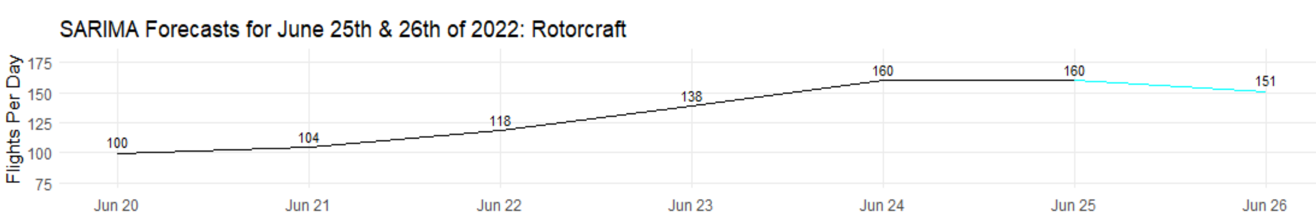
Appendix E: Extra Time-Series Forecasts for Specific Dates

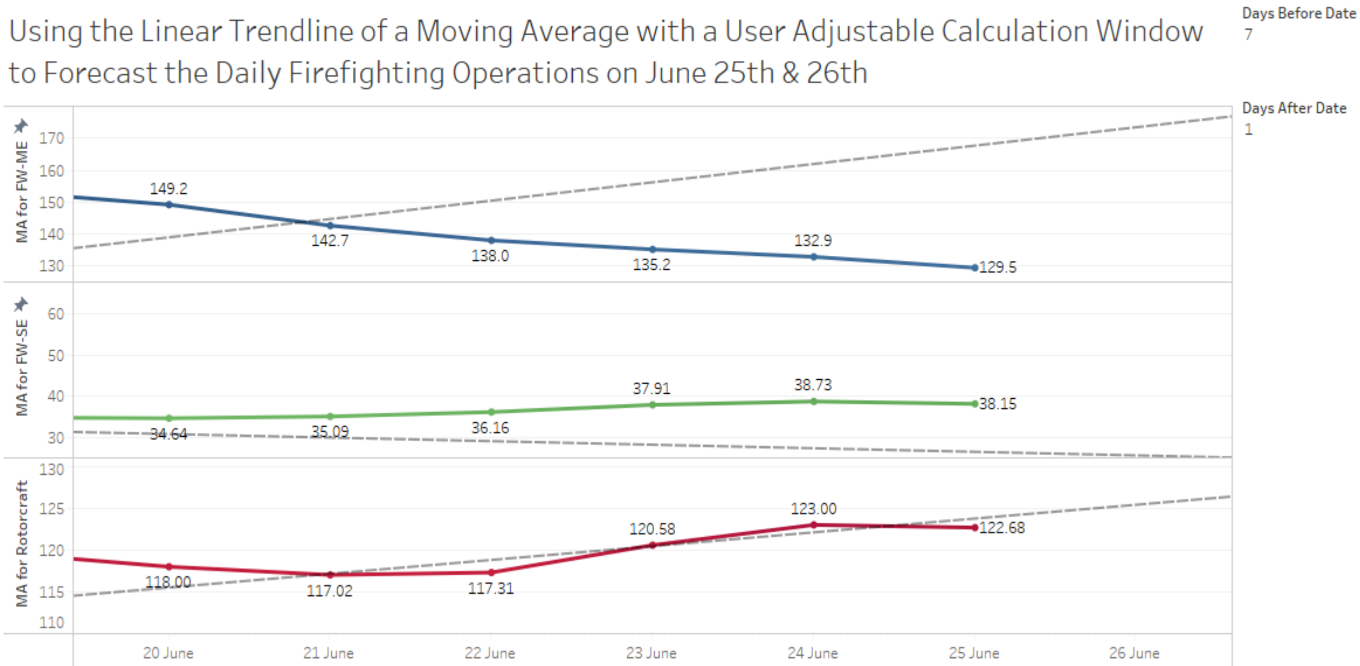
Part 1 - Line charts showing the predictions made by SARIMA, Exponential Smoothing, and Moving Average for June 25th & 26th, July 3rd & 4th, July 16th & 17th, January 11th & 12th, March 1st & 2nd and then their actual values on those days.
The ES, MA, SARIMA, and actual charts for June 25th & 26th 2022:

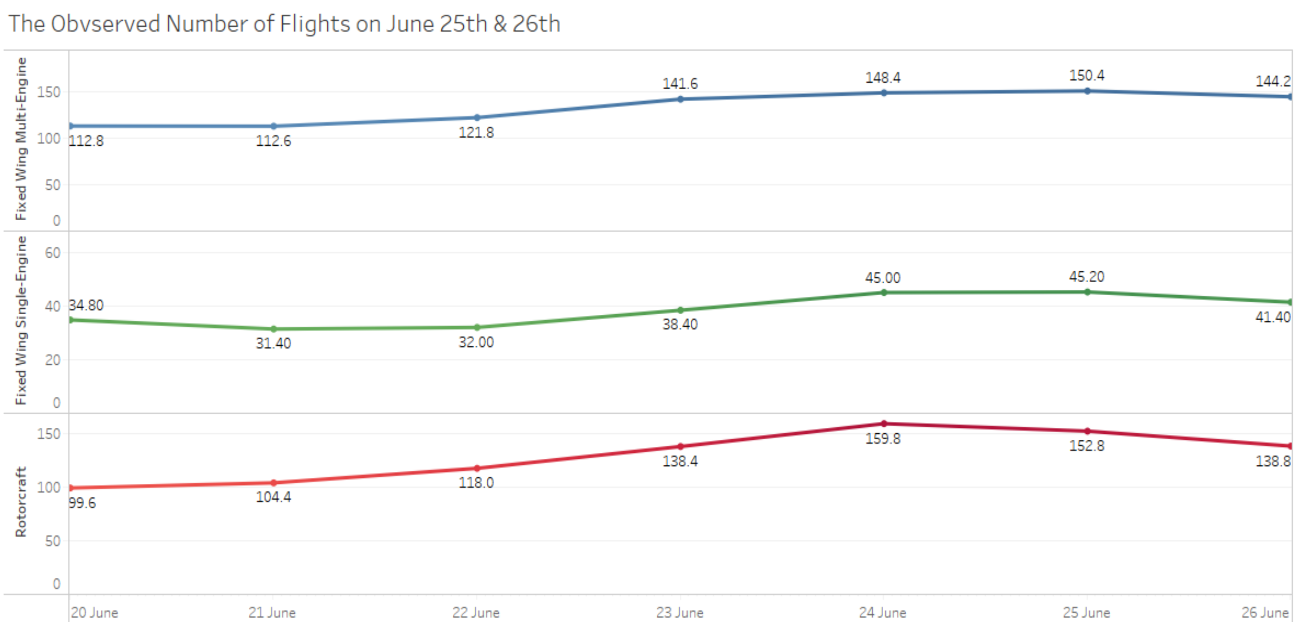








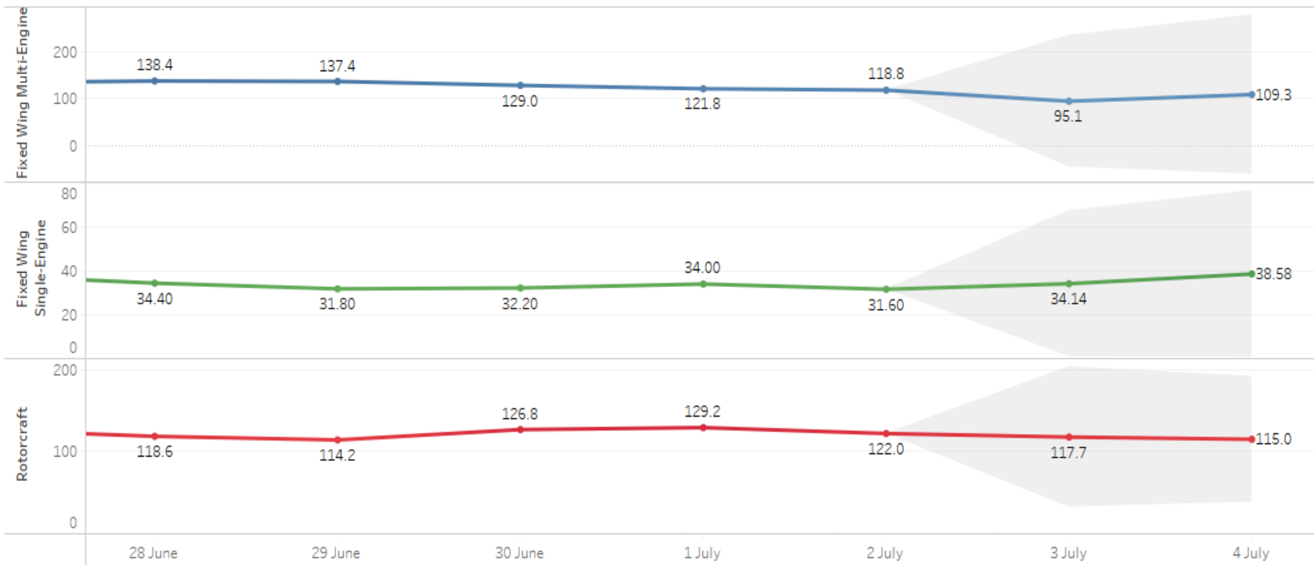




SARIMA makes the most accurate forecasts for June 25th & 26th.

The same 4 charts for July 3rd & 4th:

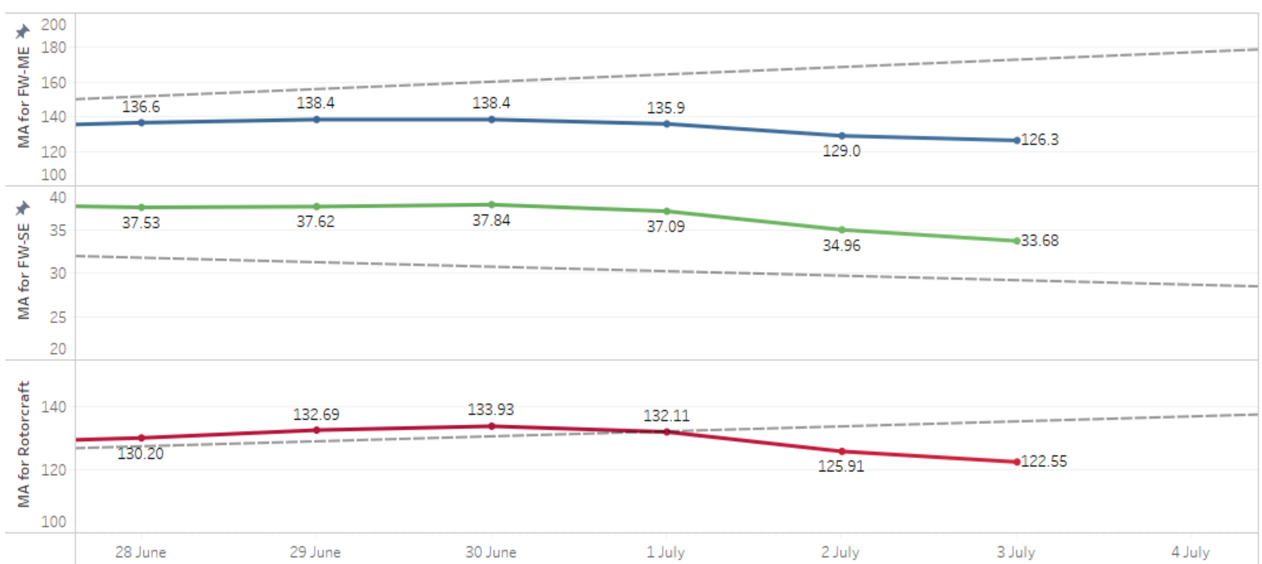
Using Exponential Smoothing with Multiplicative Seasonal Effects to Forecast Daily Firefighting Operations, by Aircraft Type, on July 3rd & 4th

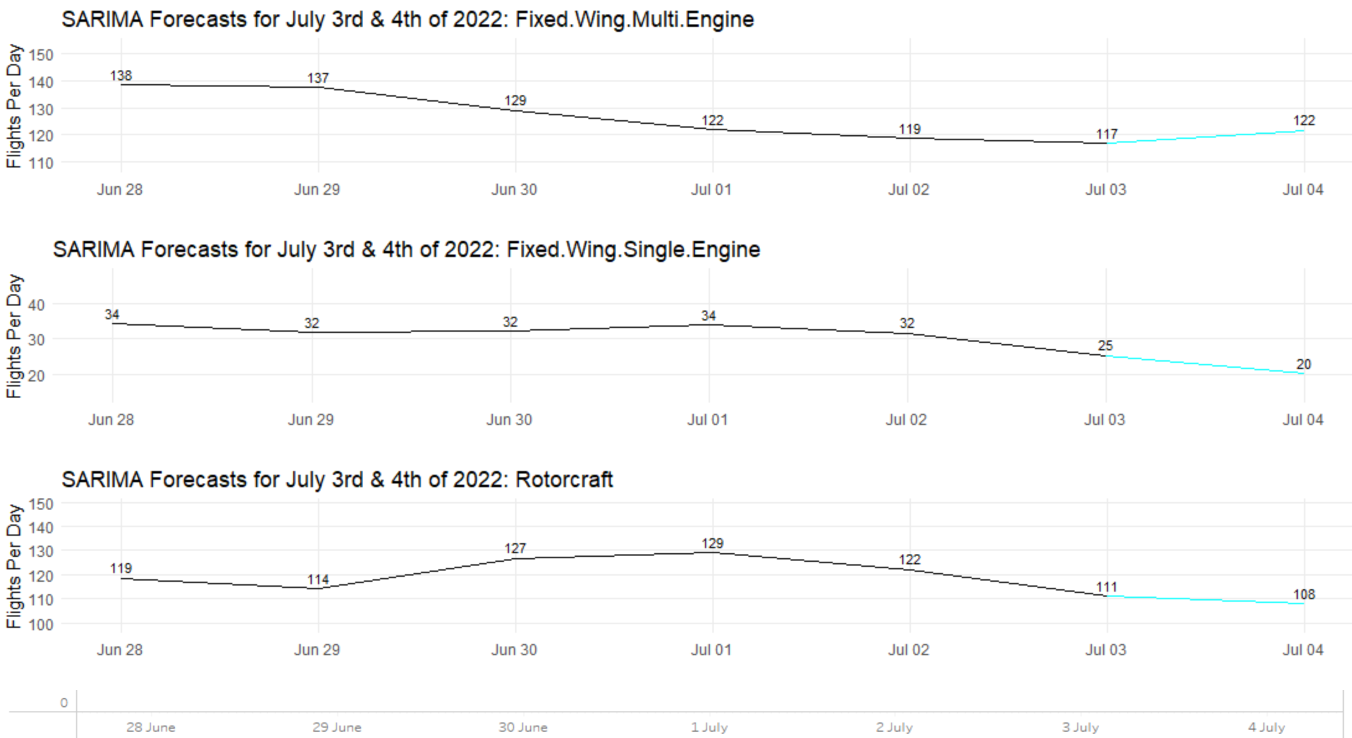


Using the Linear Trendline of a Moving Average with a User Adjustable Calculation Window to Forecast the Daily Firefighting Operations on July 3rd & 4th of 2022

Days Before Date
7

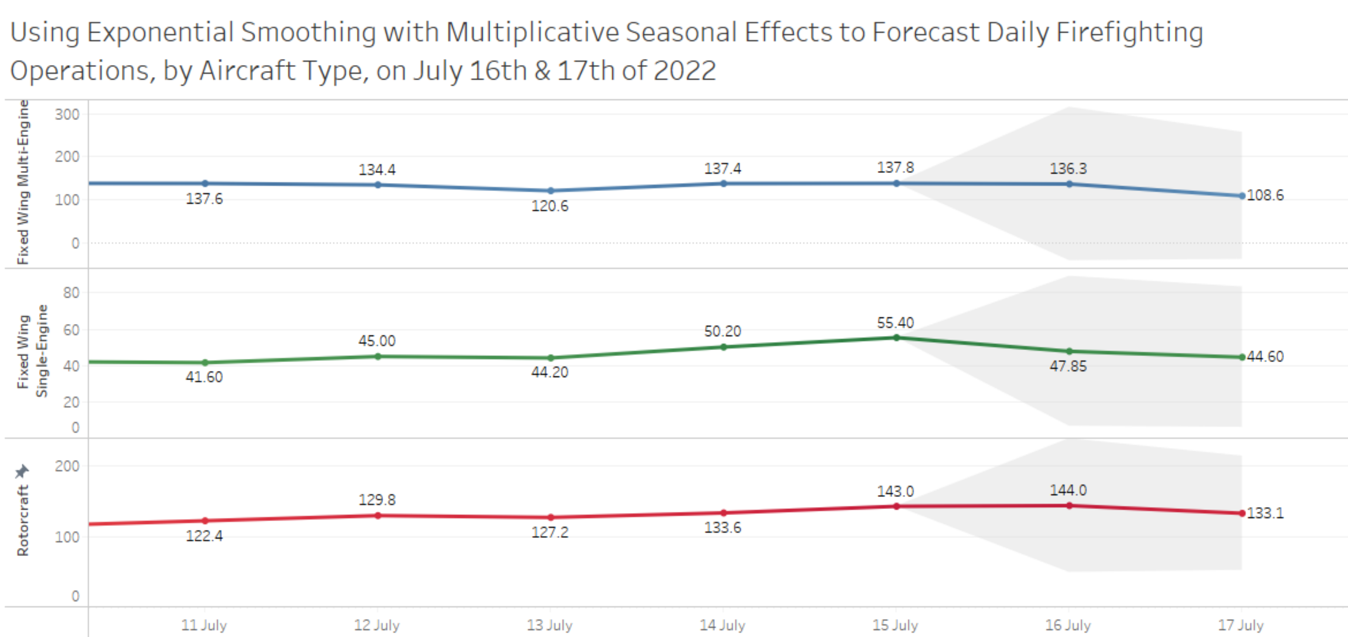
Days After Date
1

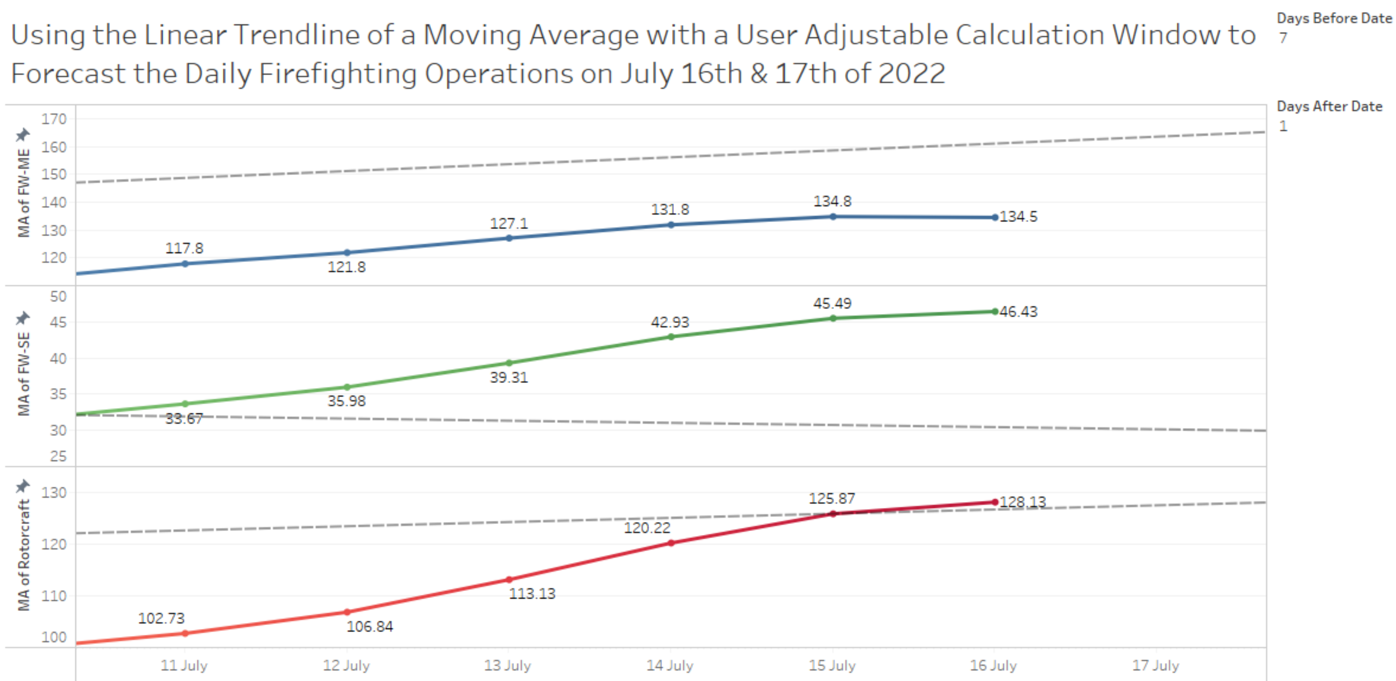


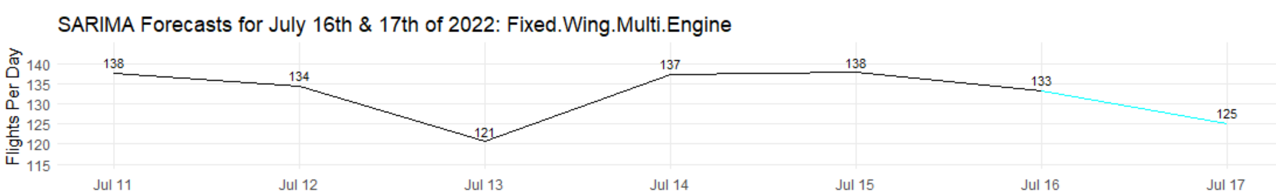


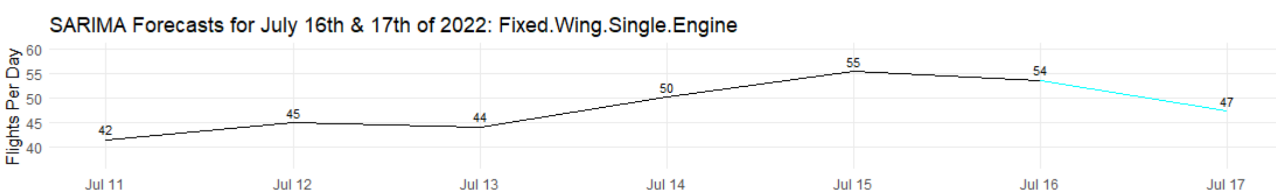
The most accurate forecasts for July 3rd & 4th are made by SARIMA.

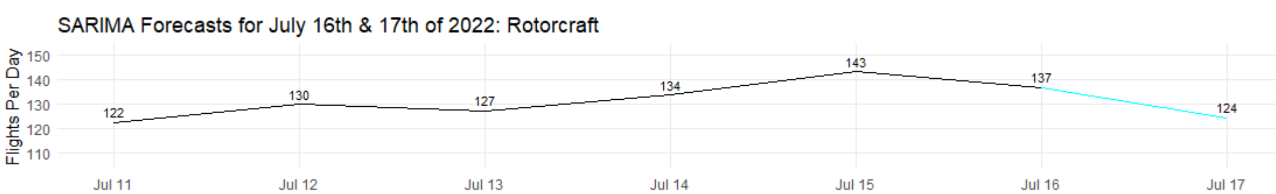
The same 4 charts forecasting the number of flights on July 16th & 17th of 2022:



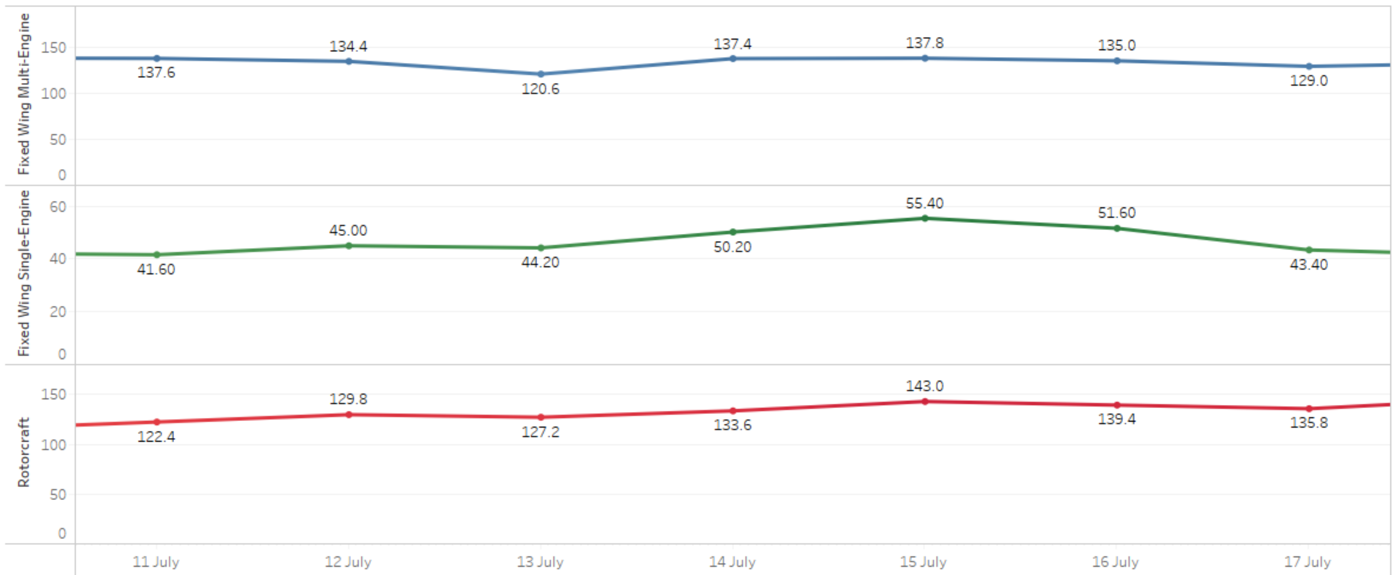

Using the Linear Trendline of a Moving Average with a User Adjustable Calculation Window to Forecast the Daily Firefighting Operations on July 16th & 17th of 2022
Days Before Date
7
Days After Date
1
MA of FW-ME
117.8
121.8
127.1
131.8
134.8
134.5
MA of FW-SE
33.67
35.98
39.31
42.93
45.49
46.43
MA of Rotorcraft
102.73
106.84
113.13
120.22
125.87
128.13
11 July
12 July
13 July
14 July
15 July
16 July
17 July


SARIMA Forecasts for July 16th & 17th of 2022: Fixed.Wing.Multi.Engine
Flights Per Day
138
134
121
137
138
133
125
Jul 11
Jul 12
Jul 13
Jul 14
Jul 15
Jul 16
Jul 17


SARIMA Forecasts for July 16th & 17th of 2022: Fixed.Wing.Single.Engine
Flights Per Day
42
45
44
50
55
54
47
Jul 11
Jul 12
Jul 13
Jul 14
Jul 15
Jul 16
Jul 17


SARIMA Forecasts for July 16th & 17th of 2022: Rotorcraft
Flights Per Day
122
130
127
134
143
137
124
Jul 11
Jul 12
Jul 13
Jul 14
Jul 15
Jul 16
Jul 17

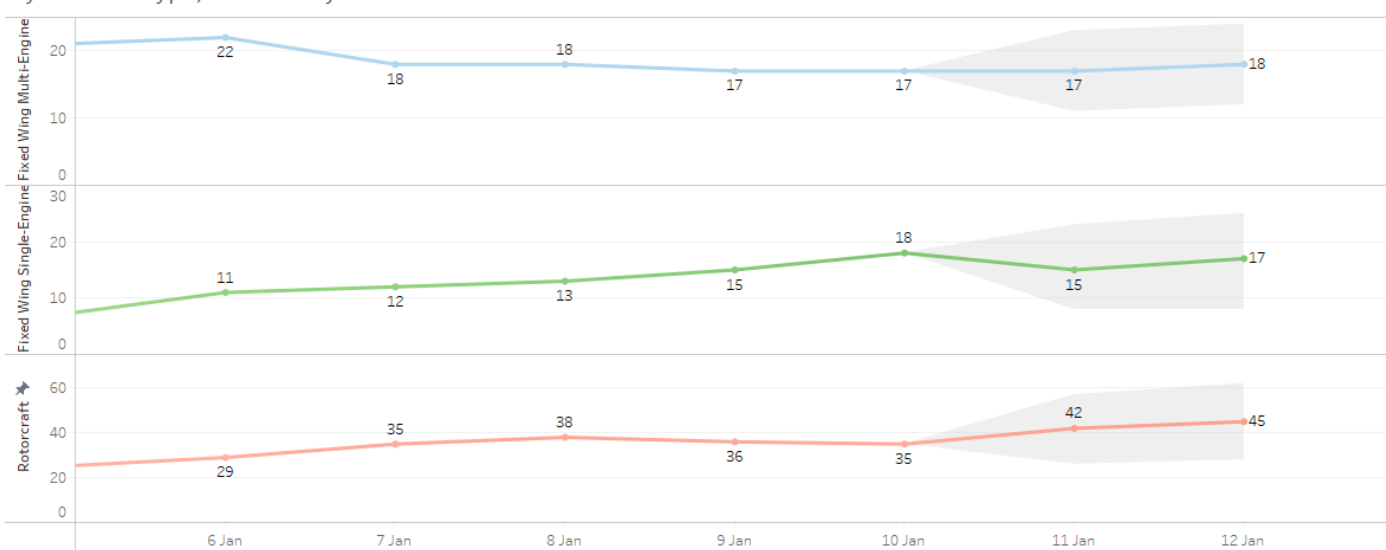
The Observed Number of Flights on July 16th & 17th of 2022

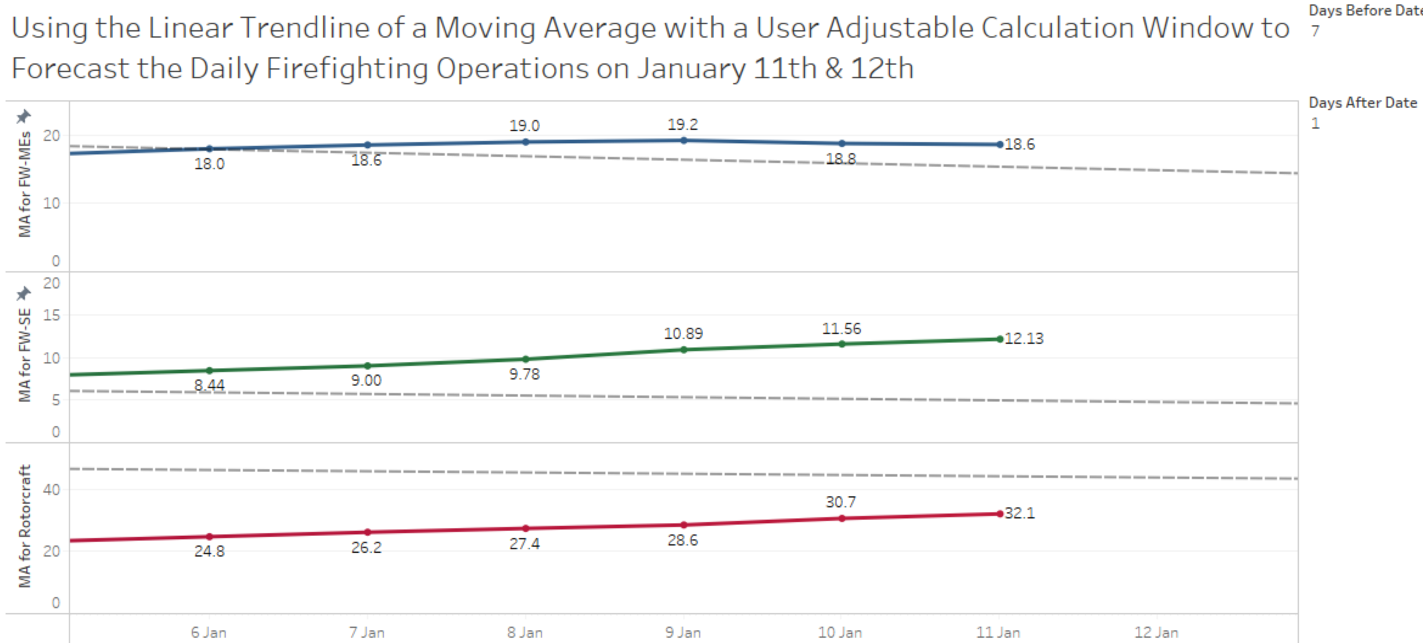


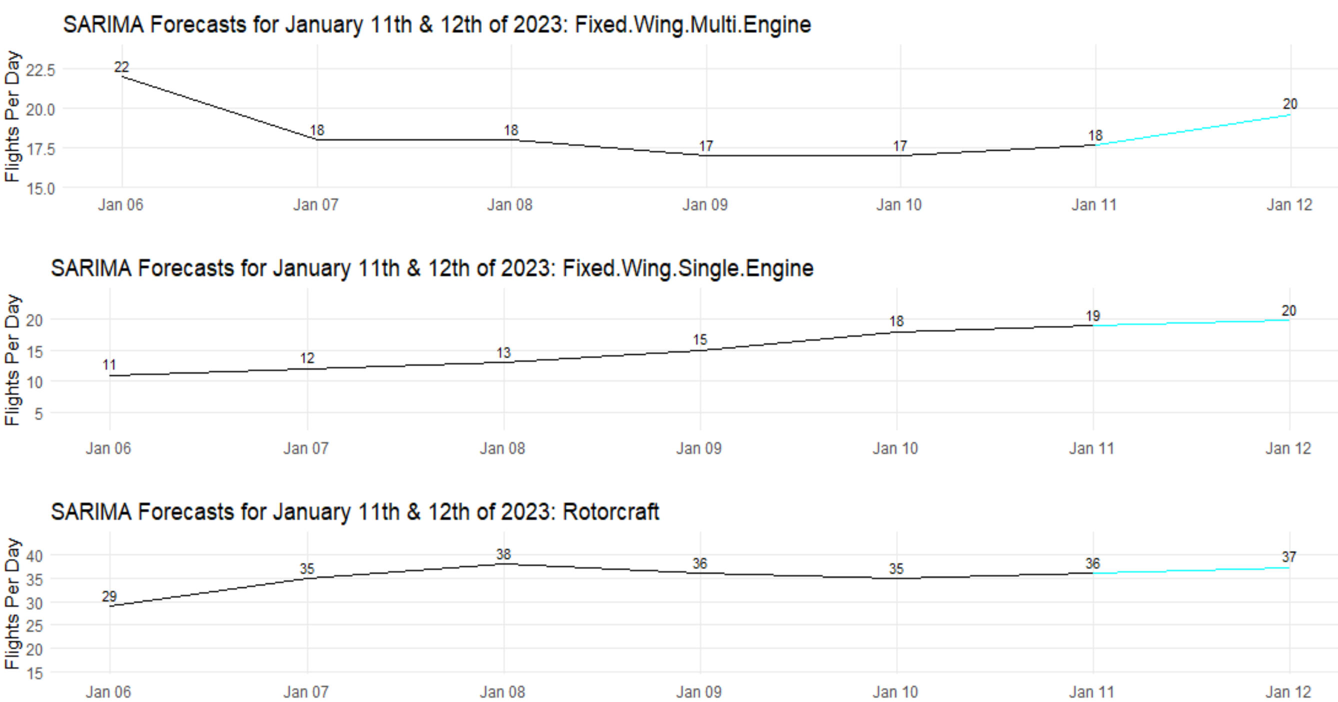
The most accurate forecasts for July 16^{th} & 17^{th} was a tie between SARIMA & $e^{smoothing}$.

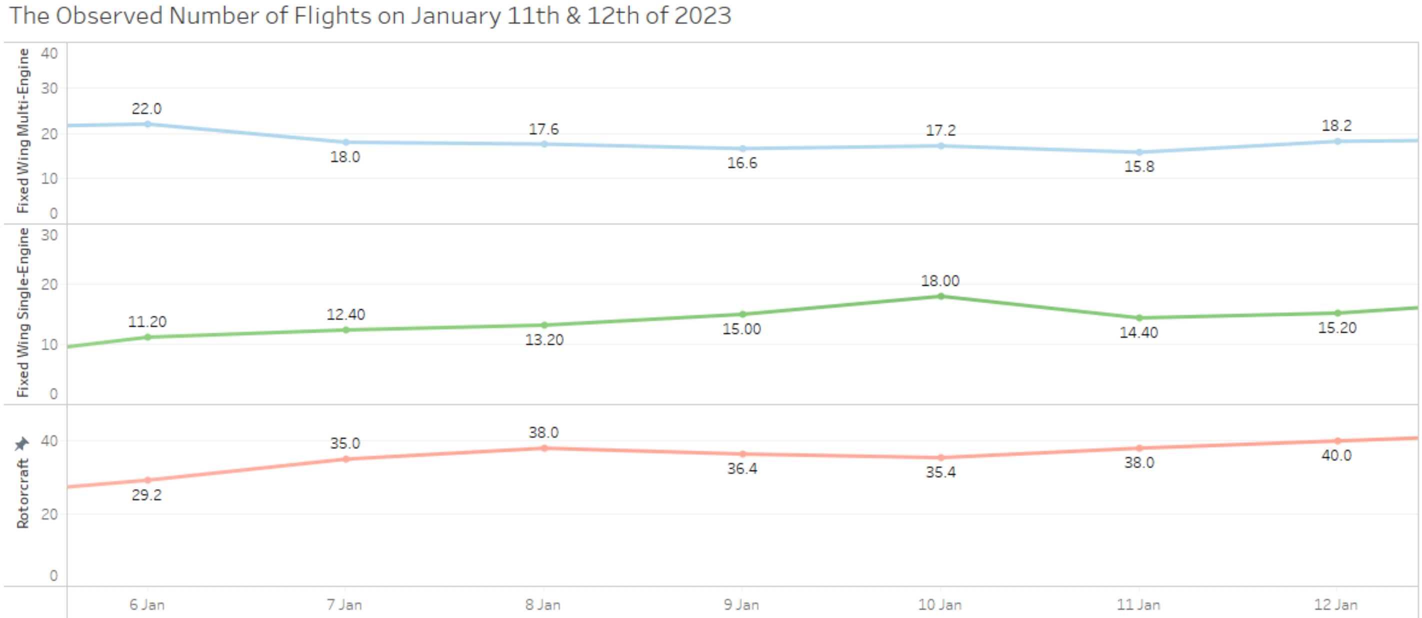
The three forecast charts and the one actual chart for January 11^{th} & 12^{th} of 2023:

Using Exponential Smoothing with Multiplicative Seasonal Effects to Forecast Daily Firefighting Operations, by Aircraft Type, on January 11th & 12th



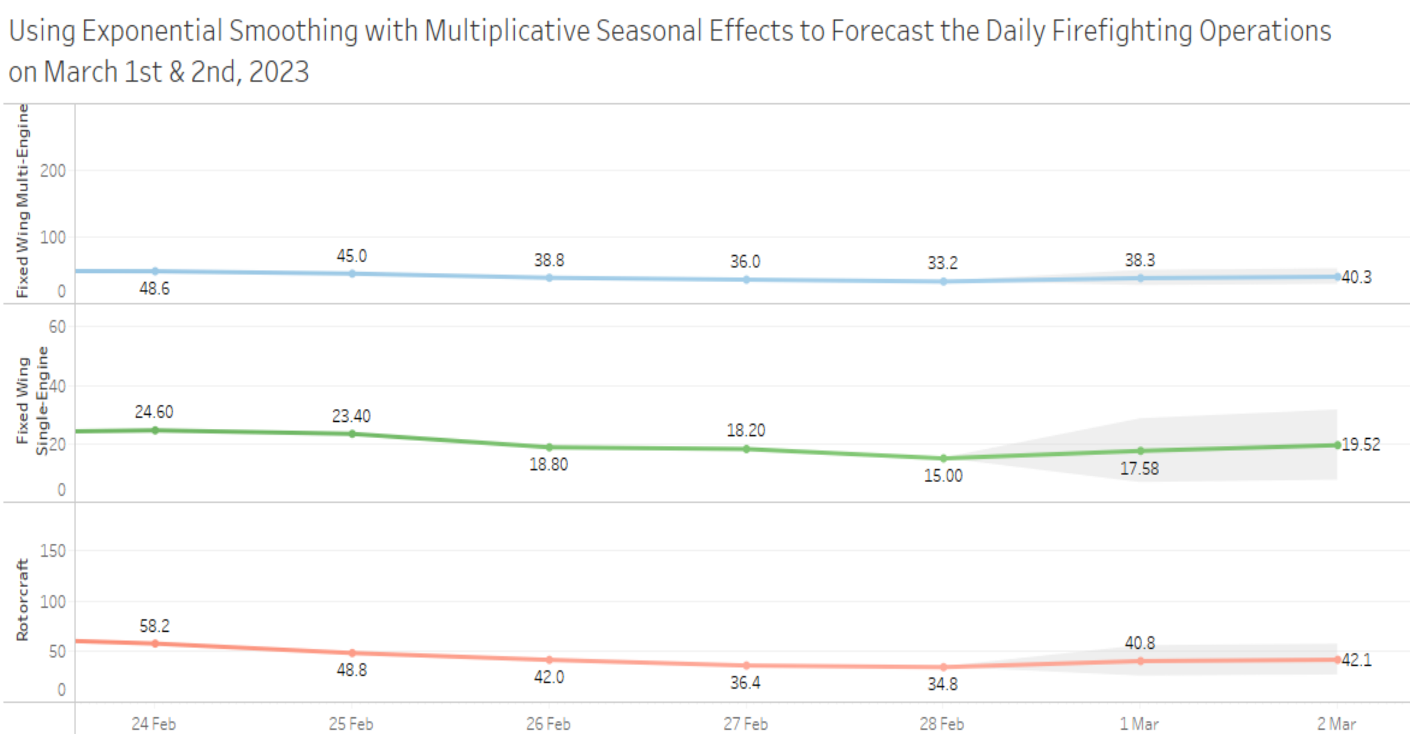

Using the Linear Trendline of a Moving Average with a User Adjustable Calculation Window to Forecast the Daily Firefighting Operations on January 11th & 12th
Days Before Date
7
Days After Date
1
MA for FW-MEs
MA for FW-SE
MA for Rotorcraft
18.0
18.6
19.0
19.2
18.8
18.6
8.44
9.00
9.78
10.89
11.56
12.13
24.8
26.2
27.4
28.6
30.7
32.1
6 Jan
7 Jan
8 Jan
9 Jan
10 Jan
11 Jan
12 Jan

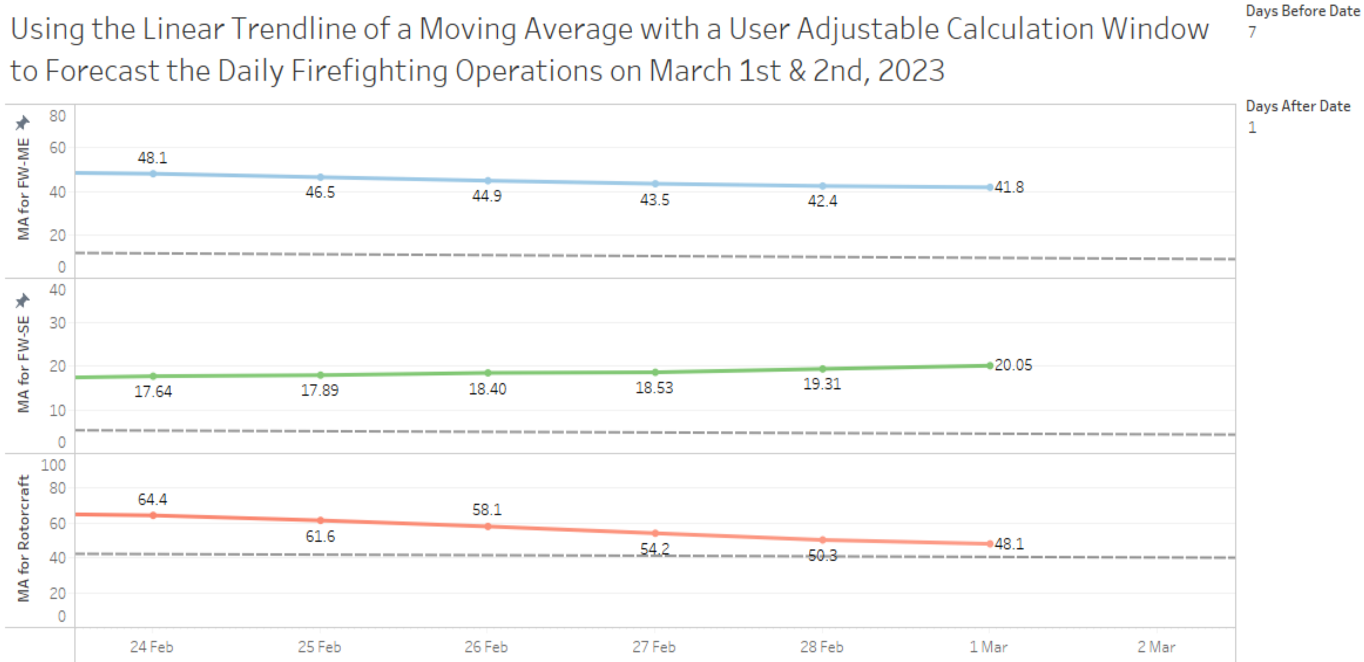

SARIMA Forecasts for January 11th & 12th of 2023: Fixed.Wing.Multi.Engine
Flights Per Day
22
18
18
17
17
18
20
SARIMA Forecasts for January 11th & 12th of 2023: Fixed.Wing.Single.Engine
Flights Per Day
11
12
13
15
18
19
20
SARIMA Forecasts for January 11th & 12th of 2023: Rotorcraft
Flights Per Day
29
35
38
36
35
36
37
Jan 06
Jan 07
Jan 08
Jan 09
Jan 10
Jan 11
Jan 12

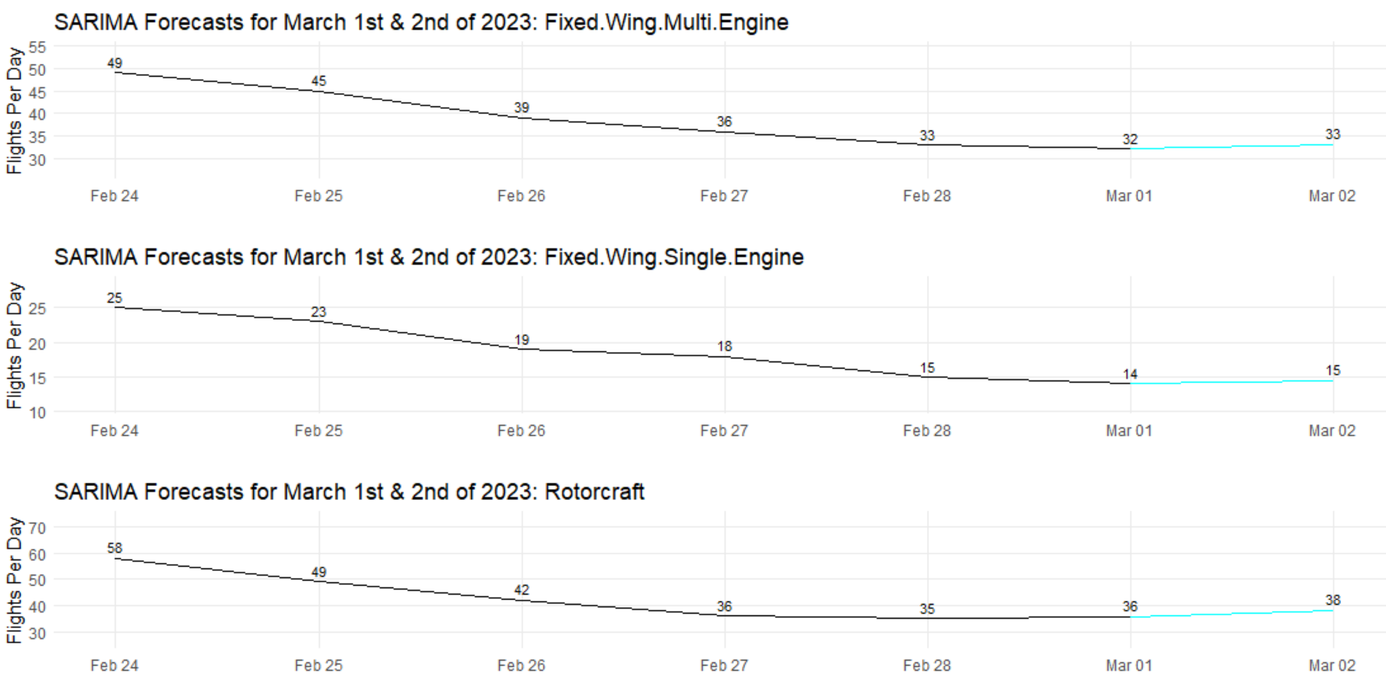


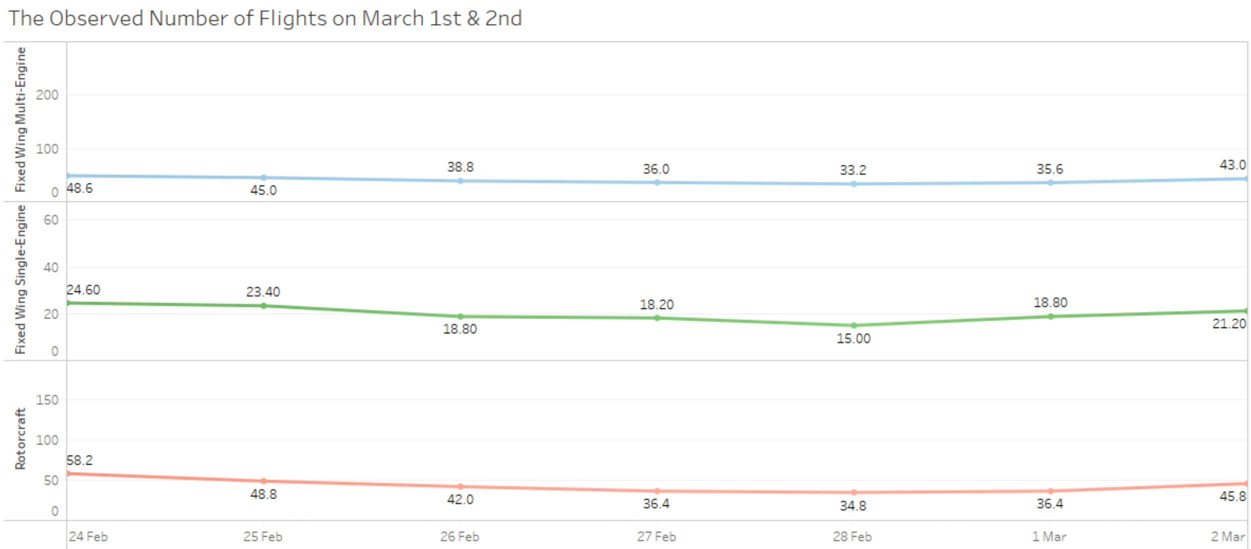
The most accurate forecasts for January 11th & 12th were made by Exponential Smoothing.

The same set of 4 charts for March 1st & 2nd of 2023:




Using the Linear Trendline of a Moving Average with a User Adjustable Calculation Window to Forecast the Daily Firefighting Operations on March 1st & 2nd, 2023
Days Before Date
7
Days After Date
1
MA for FW-ME
48.1
46.5
44.9
43.5
42.4
41.8
MA for FW-SE
17.64
17.89
18.40
18.53
19.31
20.05
MA for Rotorcraft
64.4
61.6
58.1
54.2
50.3
48.1
24 Feb
25 Feb
26 Feb
27 Feb
28 Feb
1 Mar
2 Mar


SARIMA Forecasts for March 1st & 2nd of 2023: Fixed.Wing.Multi.Engine
Flights Per Day
49
45
39
36
33
32
33
SARIMA Forecasts for March 1st & 2nd of 2023: Fixed.Wing.Single.Engine
Flights Per Day
25
23
19
18
15
14
15
SARIMA Forecasts for March 1st & 2nd of 2023: Rotorcraft
Flights Per Day
58
49
42
36
35
36
38
Feb 24
Feb 25
Feb 26
Feb 27
Feb 28
Mar 01
Mar 02



The most accurate forecasts for March 1st & 2nd of 2023 were made by Exponential Smoothing.

An Excel Table was created to enable quantitative comparisons of SARIMA, Exponential Smoothing, and Moving Average performance which has only the raw observed values for all selected dates and the integer rounded values forecasted for them by each forecasting model used rather than having the percentage distances between each forecast and the corresponding observed values which is what the table included in the body of the report conveys. Here is that table:

Date	Actual			SARIMA			Exponential Smoothing			Moving Average		
Type	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**
25/6/22	150	45	153	139	44	160	140	47	115	170	27	124
26/6/22	144	41	139	133	38	151	105	49	109	177	26	128
3/7/22	86	26	104	117	25	111	113	33	124	174	31	137
4/7/22	85	23	90	122	20	108	106	33	124	179	30	143
16/7/22	135	52	139	133	54	137	136	48	144	161	32	127
17/7/22	129	43	136	125	47	124	109	45	133	164	31	129
19/8/22	138	37	143	150	44	154	130	41	135	142	41	141
20/8/22	125	36	136	151	41	151	130	41	135	144	41	144
7/9/22	252	34	162	245	35	155	219	35	142	152	39	135
8/9/22	282	36	186	261	38	163	221	38	151	152	39	136
11/1/23	16	14	38	18	19	36	17	15	42	16	5	43
12/1/23	18	15	40	20	20	37	18	17	45	14	4	42

1/3/23	36	19	36	32	14	36	38	18	41	9	4	40
2/3/23	43	21	46	33	15	38	40	20	42	8	4	39

And there was also an intermediate table created in between the table above and the one included at the tail end of sub section 4.3 in the body of this report, namely, a table with the raw numerical difference between the forecasted # of flights per day and the actual number. This is that table:

Date	**Actual**			**SARIMA**			**Exponential Smoothing**			**Moving Average**		
Type	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**	**FW-ME**	**FW-SE**	**Rotor**
25/6/22	150	45	153	-11	-1	7	-10	2	-38	20	-18	-29
26/6/22	144	41	139	-11	-3	12	-39	8	-30	33	-15	-11
3/7/22	86	26	104	31	-1	7	27	7	20	88	5	33
4/7/22	85	23	90	37	-3	18	21	10	34	94	7	53
16/7/22	135	52	139	-2	2	-2	1	-4	5	26	-20	-12
17/7/22	129	43	136	-4	4	-12	-20	2	-3	35	-12	-7
19/8/22	138	37	143	12	7	11	-8	4	-8	4	4	-2
20/8/22	125	36	136	26	5	15	5	5	-1	19	5	8
7/9/22	252	34	162	-7	1	-7	-33	1	-20	-100	5	-27
8/9/22	282	36	186	-21	2	-23	-61	2	-35	-130	3	-50
11/1/23	16	14	38	2	5	-2	-1	-4	6	-1	-10	1
12/1/23	18	15	40	2	5	-3	-2	-3	8	-4	-13	-3
1/3/23	36	19	36	-4	-5	0	6	4	5	-29	-14	-1
2/3/23	43	21	46	-10	-6	-8	7	5	4	-32	-16	-3

References

Works Cited

2020 Aviation Annual Report. (2021). U.S. Forest Service. Retrieved February 9, 2023, from https://www.fs.usda.gov/sites/default/files/2021-06/CY2020_USFSAviationReport_Final_1.pdf

2021 Aviation Annual Report. (2022). U.S. Forest Service. Retrieved February 9, 2023, from https://www.fs.usda.gov/sites/default/files/2022-11/CY21-Aviation-Report.pdf

Amazon AWS. (2023). *AmazonS3*. Retrieved from https://docs.aws.amazon.com/AmazonS3/latest/userguide/UsingBucket.html

Borrelli, L. (2022). *U.S. Wildfire Statistics*. Bankrate. https://www.bankrate.com/insurance/homeowners-insurance/wildfire-statistics/

Fairview Fire now 94% contained; all evacuation orders and warnings lifted. (2022). CBS News. https://www.cbsnews.com/losangeles/live-updates/fairview-fire-now-62-contained-many-residents-can-return-home/

Federal Excess Personal Property Program & Firefighter Property Program. US Forest Service. (n.d.). Retrieved May 6, 2023, from https://www.fs.usda.gov/managing-land/fire/fepp

Fire Aviation Guidelines. (2014). International Fire Aviation Working Group. http://www.ifawg.net/wp-content/uploads/2017/08/IFAWG-Fire-Aviation-Guidelines-Part-1-Edition-1.0.pdf

FireLine® State Risk Report- California. (2021). Verisk Analytics. Retrieved February 13, 2023, from https://www.verisk.com/siteassets/media/campaigns/gated/underwriting/fireline-state-risk-report/california.pdf

Halpern, N., & Bråthen, S. (2011). Impact of Airports on Regional Accessibility and Social Development. Journal of Transport Geography, 19(6), 1145–1154. https://doi.org/10.1016/j.jtrangeo.2010.11.006

Modular Airborne Fire Fighting Systems (MAFFS). (2017). US Forest Service. https://www.fs.usda.gov/managing-land/fire/planes/maffs

Pierobon, M. (2021, September 14). Aerial firefighting technology and Tactics. AirMed&Rescue. Retrieved May 6, 2023, from https://www.airmedandrescue.com/latest/long-read/aerial-firefighting-technology-and-tactics

Planes | US Forest Service. (2023). www.fs.usda.gov. https://www.fs.usda.gov/managing-land/fire/planes

Study on Wildfire Fighting Resources Sharing Models Final Report. (2010). GHK Consulting. https://gfmc.online/wp-content/uploads/EC-Study-Aerial-Wildfire-Fighting-Resources-Sharing-Models-2010.pdf

Taylor, B. N., Rasnake, N., McNutt, K., McKnight, C. L., & Daley, B. J. (2018). Rapid Ground Transport of Trauma Patients: A Moderate Distance from Trauma Center Improves Survival. Journal of Surgical Research, 232, 318–324. https://doi.org/10.1016/j.jss.2018.06.055

Wildfire Statistics. (2021). Congressional Research Service. https://sgp.fas.org/crs/misc/IF10244.pdf