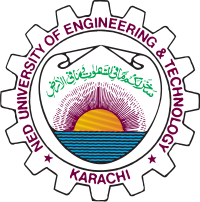
**B.E. (CS) PROJECT REPORT**

**by**

**Shaheer Shahab**



**Department of Computer and Information Systems Engineering**

**NED University of Engg. & Tech., Karachi-75270**

**B.E. (CS) PROJECT REPORT**

**Project Group:**

Shaheer Shahab CS-087

Muhammad Fahad CS-109

**BATCH:** 2014-15

**Project Advisor(s):**

Dr. Syed Abbas Ali (Internal Advisor)

Sarosh Haider (External Advisor)

**September** 2018

**Department of Computer and Information Systems Engineering**

**NED University of Engineering and Technology**

**Karachi-75270**

**ABSTRACT**

*Air Transport industry acts as a catalyst to the economic and social development of a nation. This industry encompasses all those activities which involve transportation of goods and people, by air. Air transport connects people; Air transport connects people, countries and cultures across the face of the globe. Additionally, it opens up a market to global players, thereby supporting trade and tourism significantly. The Air transport industry has contributed significantly to the growth of commerce, communication, trade and tourism globally. The popularity of travelling by airplanes is constantly growing. Much of existing research describes the global flight market. At the same time, air market is characterized by its peculiarities that have to be identified to build proper models of airspace operations. The objective of this project is to analyze air transportation market and compare the behavior of prices on flights. Using these data, an empirical data-driven model was built for air prices estimation for crossing of particular flight over particular country. There are hundreds of airlines in the world, and it is the desire of every airline to be at the top, beat all the competitors and make its reputation in airline market. Airline industries produce GBs of TBs of data per day. This data can be used in different ways to optimize their schedule, flights, services and in many areas. This data is not only used in Aviation companies but also in Shipping and Courier companies, Petroleum companies, Cab companies, Hotel and Tourism companies and many more. AI can be applied on this data to analyze, predict and to optimize work. Data mining is used to extract useful and required information from the huge pile of data. This information is used by Aviation companies to minimize their losses, by Shipping and Courier companies to optimize their work by scheduling it according to the location of the particular flight, by Petroleum companies to predict the fuel required by the airlines and many more. This information can also be used to produce an analytic report based on different factors and attributes which helps in finding the loss and profit gain by the company, what factors are causing problems and what factors needs to be improved. This report is also used to find how to get maximum profit by improving those factors which are causing problems. Thus, by applying AI on this huge data provides some useful results that helps in managing work with respect to time, helps in making better decisions, and allow to make plans which helps to improves the performance and reputation of the industry.*

**ACKNOWLEDGEMENTS**

Starting with the name of Almighty Allah, His mercy and grace. Without the will of Allah this project could never have been a possibility. We are grateful to the prayers and well wishes of our parents that have always been with us in the evens and odds, our friends who shared the good and the bad times with us and helped us put ourselves up in difficult times.

We would like to extend our heartiest gratitude to **Dr. Syed Abbas Ali** (Internal Advisor), Assistant Professor, Department of Computer & Information Systems Engineering who at every step guided us with his professional expertise and most of all we are very thankful for the confidence he put in us. Without his continuous support and assistance this project could have never been a possibility. The project being of research nature, her knowledge and experience became our greatest asset. We thank her for her guidance and endless support. Without her undying support and motivation, it was impossible for us to reach a successful conclusion.

We would like to thank our co-internal advisor, **Ma’am Urooj Ainuddin**, Assistant Professor, Department of Computer & Information Systems Engineering for taking out time from his busy schedule to guide us throughout the project and encouraging us to cope up with the problems that we had to face during the course of project.

We are also very thankful to our external advisor **Mr. Sarosh Haider**, CTO MRI, being a source of constant support and encouragement and providing us help in resources (ADSB-EXCHANGE API) for performing competitive analysis.

The High-Performance Computing Centre (HPCC), at NED University of Engineering and Technology, supported this research work too in its initial stage, so we are thankful to **Sir Ali Ismail** as well.

We are the persons who are just developers of this project, any project related ethical issues, sue issues is not our responsibility.

**CONTENTS**

**S.No. Pg No.**

**1. Introduction** 1

* 1. Background 1
  2. Motivation and need 2
  3. Objective 3
     1. Project Beneficiaries
     2. Project Applications 3
  4. Methodology
     1. Fuel consumed calculation
     2. Charge estimation
     3. Estimation elapsed time and trace route
     4. Comparison of flights
     5. Final report generation 5
  5. Block diagram of project 5

**2. Literature Review**  6

2.1. Literature review of each module 5

2.1.1. Fuel economy 6

2.1.2. Charge economy 8

2.1.3 If Delay Cost

2.1.4 Competitive analysis need

**3. Spotlight: Deep dive into Data Analytics**

3.1. Footsteps of Data analytics 19

3.2 Characteristics of Data analyst 20

3.3 Jobs related to Data analytics 23

**4. Hello World!**

**8. Conclusion** 98

**Appendix A: Name of First Appendix** 100

**Appendix B: Name of Second Appendix** 103

**Appendix C: Name of Third Appendix** 105

**References** 106

**Bibliography** 107

**CHAPTER 1**

**Introduction**

Within the few last years, the aviation market has developed and changed significantly. Growing globalization has connected the markets more and more to each other, but also the growing population has led to an increased demand in flight travel. The extent of competition in the airline industry is high, as there are many airlines competing for their reputation in the industry. To determine the profitability of the airline industry, we have performed an industry competitive analysis. This industry analysis will help us in understanding the size of the Potential Industry Earnings (PIE), and how much of this the different participants can extract.

* 1. **Background**

There is intense rivalry among different airlines. In the pre-deregulation days, airlines competed mostly on things like service, meals and in-flight movies etc., since prices were mandated by the Civil Aeronautics Board. In the post-de-regulation era, this rivalry has taken on the form of severe price competition, with airlines ruthlessly undercutting each other with reputation characteristics metrics. There are a number of airlines making the airline industry fairly crowded. Even though the 3-firm concentration in 1992 was 50%, and the 8-firm concentration was 92%, the fact that the airlines competed on price made the industry much more competitive than the numbers might suggested.

The competition is extremely intense and each competitor attempts to gain dominance over others. Besides, one can wonder how one can differentiates itself to be better than others.

This project investigates how firms build competitive advantage by focusing on the actions and responses of rivals in the airline industry. We tested the importance of response (information) as it relates to competitive advantage by linking response with performance. Findings supported the notion that a firm’s response can be analyzed from the manner in which it interprets and processes and produce meaningful results.

* 1. **Motivation and Need**

Over the past 10 years, airline travel has become the main means of transportation. Like most industries, the airline industry is impacted by the economic cycle peaks and troughs. In fact, statistics show that there is a positive correlation between GDP and air revenue demand variable. The demand for passenger travel increased significantly since 2014, explained by the economic growth and low oil prices. As a result, the economic growth rises business confidence, industrial production, and trade, which leads to rapid demand for passenger travel. However, this growth in fact hurts the industry on the ground because of computing systems, runways, airport gates, etc. In addition to that, security factor, extreme weather conditions, and national aviation system (NAS) also slow operations of the system. Together,

they lead to massive flight delays. We will make such an engine that will estimates the fuel consumption while covering its distance from one country to another. It will estimate the charge applied to it in case of entering in area of another country, trace the route of flight by user input, also interested user can identify the placement of plane any time and can be aware of delay, also airline can sum out the fuel consumed and will optimize its operations. The continuous increase of storage capacities and computational power is currently pulling the development of data analytics. Indeed, companies (and especially IT-intensive ones) are collecting massive volume of data (often referred as Big Data), such as web logs, customer information, production and sales tracking, etc.

However, the complexity in flight patterns allows us to use multiple factors in historical data to suggest the likelihood of a flight operations optimization, Airliners tend to avoid publishing their negative performance results as the airline industry is highly competitive. For this, our team wants to estimate a number of operations for its optimization.

The proposed system will enable airliners to have complete detailed report on competitive characteristics metrics that will help them making pre-plans and better decision making.

* 1. **Objectives:**
* To anticipate and deals with charges airline have to give.
* To estimate fuel consumption and many other factors.
* To develop a business model to reduce effort of particular airline.
* To examine causes for flight delays.
* To optimize flight operations.
* To reduce further economic loss for airlines.
* To lessen inconvenience occurred to people at airport.

**1.3.1 Project Beneficiaries:**

Competitive advantage refers to the strategic advantage that a firm has over its competitors. Attaining a competitive advantage reinforces a firm placing it in a prime position within its business environment. Competitive advantage is realized, when a firm gains an aspect or a collection of aspects permitting it to do better than its rivals. These aspects may comprise of right to use natural resources, for example, high quality ores or cheap power; or possession of highly skilled human resources. Recent advancements in information technology have also led to emergence of new types of competitive advantages. The internet, for instance, has enabled companies to develop unique information-related products that take the form of web content and software. It has also led to elimination of middlemen and their related costs by allowing business to transmit information directly to customers. Since time immemorial, the role of middlemen has been to bridge the information gap between consumers and firms a function which has now being seized by the internet. Because of this ability of the internet, it is now possible for a company to develop a competitive advantage by establishing an effective website. The expression competitive advantage means the capability acquired through characteristics and resources to outperform competitors in the same sector or market. There has been increased research interest regarding the sources of competitive advantage for firms, following sustained superior performance by various companies around the globe. A firm is regarded as having a competitive

advantage, if it is carrying out a value creating strategy that is not presently being executed by any other existing or potential players in the industry (Barney 1991). Successful execution of such strategies raises a company to exceptional performance by enabling the firm with a competitive advantage to do better than current or potential competitors. Competitive advantage, therefore, implies the ability to outperform current and potential rivals. Such a superior performance attained through possession of a competitive advantage helps a company to achieve and maintain market leadership. In addition, resources possessed by a company and its business strategy have a deep effect on the production of a competitive advantage. Porter (1980) sees business strategy as a mechanism that plays around with resources to produce competitive advantage. Thus, a feasible business strategy may not be sufficient to achieve a competitive advantage, if the firm does not possess exclusive resources that can generate a unique advantage. Competitive advantage promotes the survival of a firm by putting it in a superior position relative to other firms in the market. Some pinpoint benefits of project are:

* Provide user with report beneficial to optimize and overcome their operation problems.
* Improve company revenue and lower costs.
* Help in decision making.
* Cut down the overall cost and create a better environment.

**1.3.2 Project Applications:**

The purpose of the competitive analysis is to determine the strengths and weaknesses of the competitors within your market, strategies that will provide you with a distinct advantage, the barriers that can be developed in order to prevent competition from entering your market, and any weaknesses that can be exploited within the product development cycle. With markets overseeing rapid change in products and competitors, engaging in the competitive intelligence of business strategies has become mandatory for companies. This is mainly because of a short span of product life cycle, globalization etc. Hence generating intelligence has become essential for survival for reputation. The main use or application of our project is none other than aviation industry, which will be beneficial in terms of analyzing the characteristics of competitor, and can investigate the need of change required. They can have complete report on performance of other company. It may also be used by Petroleum industry for the sake to know how much fuel they will be needing in order to deliver particular flight by accessing their fuel consumption.

* 1. **Methodology:**

Twentieth-century airplanes generated a lot of data – about the engine systems, fuel use etc. Today, through thousands of sensors and sophisticated digitized systems, the newest generation of jets collects exponentially more, with each flight generating more than 30 times the amount of data the previous generation of wide-bodied jets produced. While currently only about one‑tenth of the global fleet is made up of these technologically advanced aircraft, in a decade more than half of it will be. By 2026, annual data generation should reach 98 billion gigabytes, according to a 2016 estimate by Oliver Wyman.

About 4000-5000 planes fly in one hour travelling over world’s airspace at an estimated average of 6 Interval, so in one day the quantity would 24 x 6 means 144, and when 144 is multiplied by no. of estimated planes so the figure would be 72,0000 data, which means a huge dataset will be available to do analysis. If we want to pause the feed of data set, we can pause and play it and can perform analysis in bunch of data specifically 10 or 20 days’ data. The picture of data is shown below:



Figure 1: View of ADSB API data

**1.4.1 Fuel consumed calculations:**

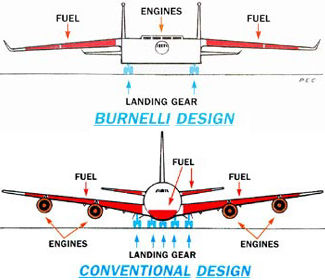
The airtransportation is today faced with the challenge of maintaining a viable economic position in the face of unparalleled increases in fuel costs in combination with ever present threat of supply shortages. Today due to high oil prices, energy crises, boomed competition, and traffic growth, fuel consumption is becoming a critical aspect in aviation industry.  In this study, fuel consumption has been considered, to be a major challenge for the air transport industry. Attributable to high oil prices and an escalation of competition, fuel consumption is rapidly becoming a critical aspect of the air transport industry. Widespread improvement in the global economy during the past year has also contributed to the demand of oil, thereby inflating its price. According to this study, world aviation oil demand was 1.18 MB/d in 1971, and reached 4.9 MB/d in 2006. The aviation sector accounts for about 5.8 % of total oil consumption worldwide. In the near future, it could be more than 45 % of all operating costs of an aircraft. The economy of a country largely depends on fuel prices. Increases in fuel consumption have an influence on the airlines in two ways; direct impact on the operating cost, and declines the demand for air travel and air cargo. Furthermore, aircraft fuel burn is proportional to CO2 emission. Therefore, as the fuel consumption increases the aviation emission shall also increase and that is a big environmental concern today. In our project, the user shall only enter the input for some flight details and will came to know about consumption of fuel per month, year, month, country, airline, person, important criteria to sum out the efficiency. Also if oil company can check that particular airlines buys this much fuel from his company, so he can estimate store required fuel in inventory. We have researched a formula for fuel consumption of airline which will discussed later on in coming chapter.

Figure 2 Fuel Placement Diagram

**1.4.2 Charge estimation:**

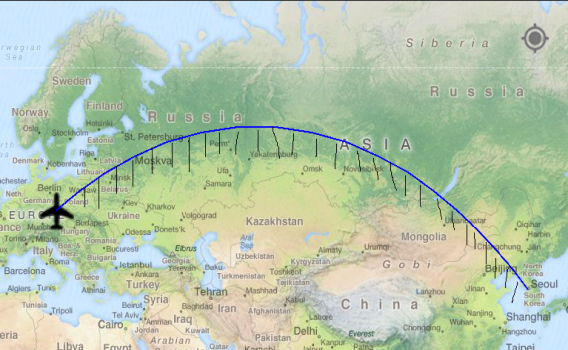
Consider any country going from one country to another hence passing by Pakistan’s (suppose) airspace, so they have to pay fine (charge) to Pakistan based on number of **miles** it covers when it enters the airspace to the time it exits and the charge per mile defined different in different countries, shown in figure, since we have data of world’s airspace but not just of Pakistan. By this project, we could be able to identify how much they would be paying charge, monthly, weekly, daily, in user defined dates, airline wise, can identify type of plane, country entering, model of plane, fuel consumed, average velocity of travel, and many other combinations. All in all, a full flesh report can be obtained by analyzing a huge data available.

Figure 3 Charge Estimation Technique through trace

**1.4.3 Estimating elapsed time to reach destination and trace route:**

There are many people, who desires to check the status of flights flying over their homes. So we are making trace route of flights. The user just has to enter some input like: arrival place, departure place, airline, their times, departure date and possible important input. At the output we would be presenting route of flight. The possible interested peoples can be: passenger on board, passenger in transit, ground staff, pilot who has to take plane for next destination, air control tower. Air hostess, ambulance, fire engine, taxi cab driver who got tag from a passenger to take him from airport. The route would be traced using long, lat, altitude from data updating every sec. If user enters arrival place and destination place, he would be getting the time (actual time + delay {if present}). When data is collected from real time, we can perform regression or correlation to predict flight delay cause by any reason. It

would be beneficial for users in sense they will be getting report of status of flight, the time they should reach to pick their dear ones. The picture of such app would be like this:

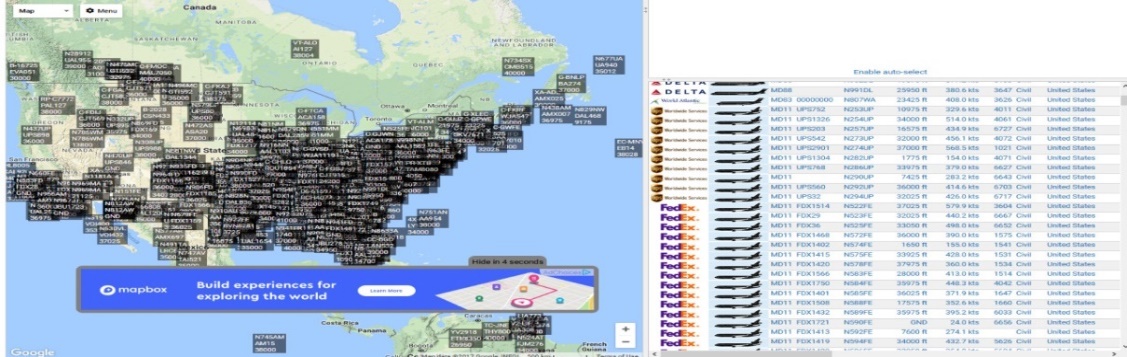


Figure 4 View of Flight tracking

**1.4.4 Comparison of flights:**

**Competitor analysis** in [marketing](https://en.wikipedia.org/wiki/Marketing) and strategic [management](https://en.wikipedia.org/wiki/Management) is an assessment of the strengths and weaknesses of current and potential [competitors](https://en.wikipedia.org/wiki/Competition_(economics)). This analysis provides both an offensive and defensive strategic context to identify opportunities and threats. Profiling combines all of the relevant sources of competitor analysis into ‘one framework in the support of efficient and effective strategy formulation, implementation, monitoring and adjustment.

Figure 5 Depicting comparison

By adding some details required as input, we can have separate cards for each airline.

**1.4.5 Final report generation:**

By having all the features analyzed respectively, now we can have complete report of what we have analyzed before in graphical way and statistical manner. One can have complete insights into owns as well as others performance characteristics and can have investigation over report to have better decision power.

Figure 6 Depicting report

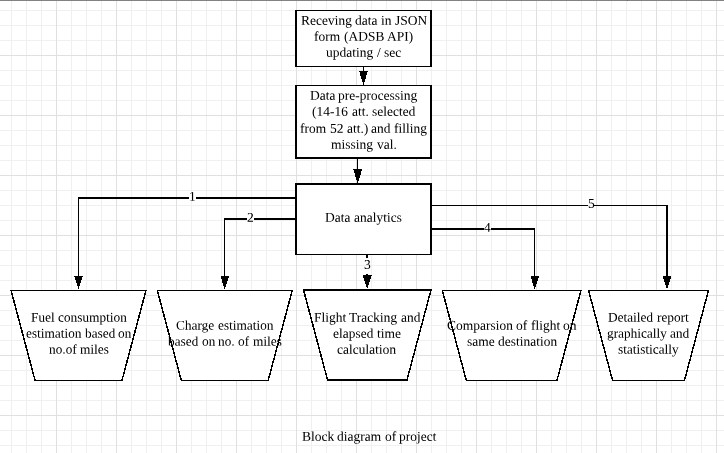
**1.5 Block diagram of Project:**

Figure 7 Block Diagram of project

**CHAPTER 2**

**Literature Review**

A **literature review** is a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic.

**2.1 Literature Review on each module:**

**2.1.1 Fuel economy:** The most significant studies on fuel consumption optimization were begun after the Arab oil embargo in 1973 when scarcity of fuel was observed. The recent decade’s research on optimization of fuel consumption has gained headway in converging upon a few important effective areas of aviation industry. The Literature proposes the technological, operational and alternate fuels and fuel property as potential areas for fuel consumption optimization. The literature reveals specific areas which are taken into account for fuel optimization includes engine design, takeoff and landing fuel properties, flight route, technological trends, etc. The recent empirical studies assent that the optimization of fuel consumption needs to develop dimensions and inter-relationship among the various parameters. But how these parameters and their sub parameter coalesce with one another, is really challenging. Thus it becomes necessary to simplify the overall structure of the methodology of optimization of fuel to make it generic and applicable for aviation industry. We examined whether optimization of fuel consumption can however be radically redesigned to accommodate various research gaps that indicated in literature with the aim to make optimization of fuel consumption more logical and productive.

The findings of the literature survey act as initial information for our next step. It describes various productive aspects of the potential areas of aviation for the optimization of fuel consumption in aviation industry.

In order to check that our study from the literature survey contains all-inclusive solutions, the initial information was clubbed together with the identified research gaps which lead to problem definition and set the objectives for our present work. It defined the problem as “engine for fuel consumption optimization in Aviation Industry”. The objectives of work as discussed above are outlined as under:

1. To explore the fuel consumption optimization in aviation industry

2. To discuss the findings on fuel consumption optimization.

This work develops informational framework of optimization of fuel consumption. The evaluation of informational framework will be done in term of reliability and validity checking. This study would form an input for aviation sector to achieve optimal fuel consumption.

**2.1.2 Charge economy:**

The restricted area exists to contain things which could cause great danger to non-participating aircraft. In short, they need to keep people out for their own safety. In practical terms, this means you cannot fly through the dimensions of a restricted airspace without getting permission from the controlling agency (usually Center). If the airspace is “active,” “open,” or “hot” they will deny you permission, and you will have to go around. You can legally fly through a restricted airspace when it is “cold,” or “closed.”

**2.1.3 Delay Cost:**

Airline industry incurs an average cost of about $11,300 per delayed flight based on 61,000 delayed flights per month average (Excludes cost to passenger and lost demand). A more accurate delay regressing system can help to verify operational variables that contribute to delays. While some conditions, such as weather, are not controllable factors, the way airlines and airports operate and optimize resources in the face of “Act of God” is controllable.

If your flight's delayed for 3 or more hours:

|  |  |  |
| --- | --- | --- |
| **Delay** | **Flight distance** | **Compensation** |
| 3 hours or more | Less than 1,500km | €250 |
|  | Between 1,500km and 3,500km | €400 |
|  | More than 1,500km and within the EU | €400 |
| 3-4 hours | More than 3,500km, between an EU and non-EU airport | €300 |

By our approach, one can have control over this delay in future, by investigating the cause of delay by real time tracking, and then comparing with others, and get to know about their strategies.

**2.1.3 Competitive analysis need:**

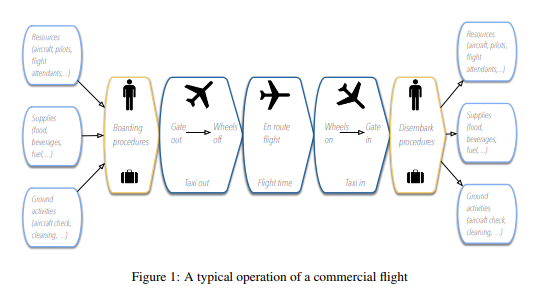
Past and present strategies of each major competitor need to be reviewed by companies. Past strategies provide insight into failures and reveal how organizations engineered changes. Reviewing competitor strategies involves the assessment of the competitor’s target market and differential advantage. It is also advantages for organizations to evaluate how successful the competitors have been in achieving its objectives and carrying out its strategies. A major objective of competitor analysis is to be able to predict competitor’s responses to market and competitive changes. Scenario planning and counter response strategies could minimize risk and assist with the implementation of own strategies and plans. Failure to understand competitors will lead to costly marketing mistakes and could have detrimental financial implications for companies. By our approach you can have complete statistics, with MIS report of the competitor one with so ease intuitively.

Figure 8 A typical operation of a commercial flight

**CHAPTER 3**

**Spotlight: Deep Dive into Data Analytics**

Data analytics is the science of collecting, organizing and analyzing very large sets of data in order to identify patterns and draw conclusions.

For example, let’s say that the executives of national fast food chain are looking to add a fresh food item to their menu. A data analyst could collect and examine data on all fast food purchases over the past 5 years. In doing so, they could identify trends and patterns in the data that would suggest the most popular fast food items being sold by competitors. This analysis can help executives make decisions about the new best food item for the menu. If examined a bit further, the data could also help executives choose specialty food items for different demographic regions.

Figure 9 Depicting data analytics

Data analytics is used in many different industries to support initiatives that are most important to an organization and to help guide critical decisions.

Real-time analysis is an emerging business tool that is changing the traditional ways enterprises do business. More and more organizations are today exploiting business analytics to enable proactive decision making; in other words, they are switching from reacting to situations to anticipating them. For most businesses and government agencies, lack of data isn’t a problem. In fact, it’s the opposite: there’s often too much information available to make a clear decision.

With so much data to sort through, you need something more from your data:

* You need to know it is the right data for answering your question.
* You need to draw accurate conclusions from that data.
* You need data that informs your decision making process

One of the reasons for the flourishing of data analytics as a tool is that it can be applied in any industry where data is captured and accessible. This data can be used for a variety of reasons, for improving the organization’s capability to offer valuable insights online and digital information. The solving of business problems using the relevant data and turning it into insights, providing the enterprise with the knowledge it needs to proactively make decisions. In this way the enterprise will gain a competitive advantage in the marketplace.

**3.1 Footsteps of Data analytics:**

To improve your data analysis skills and simplify your decisions, execute these five steps in your data analysis process:

**Step 1: Elucidate your Question:**

In your organizational or business data analysis, you must begin with the right question(s). Questions should be measurable, clear and concise. Design your questions to either qualify or disqualify potential solutions. For eg. A government contractor is experiencing rising costs and is no longer able to submit competitive contract proposals. So question cam arise is:  Can the company reduce its staff without compromising quality?

**Step 2: Set Clear Measurement Priorities:**

This step breaks down into two sub-steps:

1. **Decide What to measure:**

Using the government contractor example, consider what kind of data you’d need to answer your key question. In this case, you’d need to know the number and cost of current staff and the percentage of time they spend on necessary business functions. In answering this question, you likely need to answer many sub-questions (e.g., Are staff currently under-utilized?). Finally, in your decision on what to measure, be sure to include any reasonable objections any stakeholders might have.

1. **Decide How to measure:**

Thinking about how you measure your data is just as important, especially before the data collection phase, because your measuring process either backs up or discredits your analysis later on. Key questions to ask for this step include:

* What is your time frame? (e.g., annual versus quarterly costs)
* What is your unit of measure? (e.g., USD versus Euro)
* What factors should be included?

**Step 3: Collect and Clean Data:**

With your question clearly defined and your measurement priorities set, now it’s time to collect your data.

* Before you collect new data, determine what information could be collected from existing databases or sources on hand.
* Determine a file storing and naming system ahead of time to help all tasked team members collaborate.
* If you need to gather data via observation or interviews, then develop an interview template.
* Keep your collected data organized in a log with collection dates and add any source notes as you go.

After collecting data, it need to be clean up. **Data** cleansing or **data cleaning** is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleansing is a valuable process that can help companies save time and increase their efficiency.

**Step 4: Analyzing data:**

After you’ve collected the right data to answer your question from Step 1, it’s time for deeper data analysis. Begin by manipulating your data in a number of different ways, such as plotting it out and finding correlations. It lets you sort and filter data by different variables and lets you calculate the mean, maximum, minimum and standard deviation of your data. As you manipulate data, you may find you have the exact data you need, but more likely, you might need to revise your original question or collect more data. Either way, this initial analysis of trends, correlations, variations and outliers helps you [focus your data analysis on better answering your question](https://www.bigskyassociates.com/blog/bid/339496/Are-You-Using-The-Right-Data-Analysis-Techniques) and any objections others might have.

**Step 5: Interpret results:**

After analyzing your data, it’s finally time to interpret your results. As you interpret your analysis, keep in mind that you cannot ever prove a hypothesis true: rather, you can only fail to reject the hypothesis. Meaning that no matter how much data you collect, chance could always interfere with your results.

As you interpret the results of your data, ask yourself these key questions:

* Does the data answer your original question? How?
* Does the data help you defend against any objections? How?
* Is there any limitation on your conclusions, any angles you haven’t considered?

If your interpretation of the data holds up under all of these questions and considerations, then you likely have come to a productive conclusion. The only remaining step is to use the results of your data analysis process to decide your best course of action.

By following these five steps in your data analysis process, you make better decisions for your business or government agency because your choices are backed by data that has been robustly collected and analyzed. With practice, your data analysis gets faster and more accurate – meaning you make better, more informed decisions to run your organization most effectively.

**3.2 Characteristics of Data analyst:**

The characteristics or traits should someone in data analytics have is fundamental knowledge of technology, programming and statistics is necessary for someone to succeed in this field, but soft skills are equally as important. Some of these includes:

1. **Curious and self-motivated:**

Constantly asking questions and persistent in finding answers.

1. **Creative and flexible:**

Able to consider new and fresh perspectives to solving problems.

1. **Strategic and collaborative:**

Recognizing the importance of working as a team to solve issues from a big-picture perspective.

1. **Effective communicator:**

Able to translate numbers into actions and tell compelling stories that leads to success.

**3.3 Jobs related to Data analytics:**

The related job titles below represent a range of options for professionals with data analytics. While some titles represent entry-level positions, others may require previous experience in the field.

* Data scientist.
* Research scientist.
* Business analyst.
* Data analyst.
* System analyst.
* Business intelligence developer.
* Business system analyst.
* Business intelligence analyst.
* Financial analyst.
* ETL developer.
* Data architect.
* Solutions architect.
* Consultant.

**CHAPTER 5**

**Research Phases**

The research work is a statement of "**why**" the study is being conducted, or the goal of the study. The goal of a study might be to identify or describe a concept or to explain or predict a situation or solution to a situation that indicates the type of study to be conducted. This research is divided into three research phases including:

***First Phase****: Feature extraction and pre-processing*

***Second Phase****: Search of technique to calculate fuel consumption*

***Third Phase:*** *Search of technique to estimate charge*

***Fourth Phase****: Research on how to track flights*

***Fifth Phase****: Research on how to compare flights*

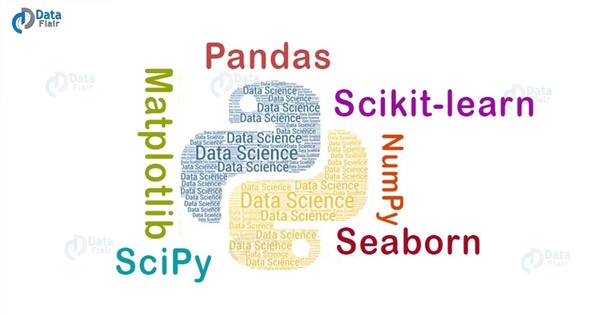
**5.1 First Phase:**

With the development of science and technology, a leap in the information age, data sets update and grow faster, number of data dimensions becomes higher and unstructured are more prominent.

large amounts of data provide utilizable information, but also make it difficult to use these data effectively. Useful knowledge may be inundated in a large number of redundant data, this will occupy a lot of storage space and computation time, and too many feature input will make the training process time-consuming, or even hinder convergence of the training network, and finally affect the precision of recognition. So how to make use of these huge volumes of data, analyze, extract useful information and exclude the influence of related or repeated factors, necessary measures should be taken to reduce the dimension of information features under the prerequisite of not affecting the problem solving as much as possible, this process is called feature extraction. Feature extraction determines the final results of the system. The basic task of feature extraction and selection is to find out a group of the most effective features for statistical representation, that is, compressing from high-dimensional feature space to low-dimensional feature space, so as to design statistics effectively.

**CHAPTER 6**

**Technologies and Libraries Used**

* 1. **Python**

When it comes to data science, data analytics is one of the significant elements used to maximize value from data. With Python as the data science tool, exploring the basics of data analytics becomes easy and effective. In a nutshell, data analytics is more about statistics, mathematical optimization, and estimation. It has become the most preferred data analytics tool in the way it allows aspirants to ‘do math’ easily.

* + 1. **Libraries used in Python**

Data science landscape is changing rapidly, and tools used for extracting value from data science have also grown in numbers. The two most popular languages that fight for the top spot are R and Python. Both are revered by enthusiasts, and both come with their strengths and weaknesses. But with the tech giants like Google showing the way to use Python and with the learning curve made short and easy, it inches ahead to become the most popular language in the data science world. The libraries used by us during project are:

 **6.1.1.1 Pandas**

Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

When you want to use Pandas for data analysis, you’ll usually use it in one of three different ways:

* Convert a Python’s list, dictionary or Numpy array to a Pandas data frame
* Open a local file using Pandas, usually a CSV file, but could also be a delimited text file (like TSV), Excel, etc
* Open a remote file or database like a CSV or a JSON on a website through a URL or read from a SQL table/database

**REFERENCES**

[1] A.B. Smith, C.D. Jones, and E.F. Roberts, “Article title with leading capitalization, and the comma inside the quotes,” *Journal (in Italics),* Month, Year

[2] Jones, C.D., A.B. Smith, and E.F. Roberts, *BOOK TITLE (in Italics, capitalized, ended with a period).* Location: Publisher, Year.