DATA ANLYTICS & VISUALIZATION

An Industrial Internship Report

Submitted in partial fulfillment for the award of the degree of

B.TECH

in

Information Technology

by

AKASH KUMAR   **17BIT0055**

(August-September) 2020

<u>DECLARATION BY THE CANDIDATE</u>

I hereby declare that the project report entitled **"Data Analytics & Visualization"** submitted by me to School of Information Technology & Engineering, Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the degree of **B.Tech (Information Technology)** is a record of bonafide **Industrial Internship − ITE3099** work carried out by me. I further declare that the work reported in this **Industrial Internship report** has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Vellore

Signature of the Candidate

Date:09/10/2019

AKASH KUMAR

School of Information Technology & Engineering [SITE]

<u>CERTIFICATE</u>

This is to certify that the Industrial Internship report entitled **"Data Analytics & Visualization"** submitted by **AKASH KUMAR (17BIT0055)** to School of Information Technology & Engineering, Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the degree of **B.Tech (Information Technology)** is a record of bonafide **Industrial Internship – ITE3099** work carried out by him in **Medulance**. The **Industrial Internship** project fulfills the requirements as per the regulations of this Institute and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Ravjot Arora

(Project Head)

7 UbBunglowRoad ,Kamla Nagar , New Delhi 110007

_____

Examiner – Panel In-Charge   **Date:09/10/2019 (Name and**
Signature)

CERTIFICATE ISSUED BY THE ORGANIZATION

MeduLance Healthcare Private Limited
Registered Address- 7 UbBunglowRoad
,Kamla Nagar , New Delhi 110007
Email- info@medulance.com
Phone- +918882978888
CIN- U74999DL2017PTC320563

Date: 15/05/2020

TO WHOM IT MAY CONCERN

This is certify that Mr. Akash Kumar (3rd year) , a student of Vellore Institute of Technology, Information Technology (I.T.) B.Tech (2017-2021) has successfully completing Data Analytics and Visualization Internship programme for the duration of 15th April , 2020 to 12th May,2020 short internship at Medulance Healthcare Private Limited. During the period of his internship he was found punctual, hardworking and inquisitive.

We wish him every success in life.

Medulance Healthcare Private Limited

Medulance Healthcare Pvt. Ltd.

(Authorized Signature)

Director

ACKNOWLEDGEMENT

I have successfully completed **"Data Analytics & Visualization"** internship in Medulance. I got to learn various new aspects and latest technology of robotics. I got to enhance my practical as well as theoretical skills in engineering throughout this internship program. It has been my privilege to have a team of project guide who have assisted me from the commencement of this project. The success of this project is a result of sheer hard work, and determination put in by me with the help of my project guide. I would also like to acknowledge and my heartfelt gratitude to Mr. Ravjot Arora (Project Head) who continuously supported me in every possible way, from initial advice to encouragement to complete my internship successfully. His wisdom, knowledge, and commitment to the highest standards inspired and motivated me. Without his insight, support, and energy, this project wouldn't have kick-started and neither would have reached fruitfulness. I express my immense pleasure and deep sense of gratitude to Mr. Ravjot Arora for giving me this opportunity and the guidance throughout the program. The project is dedicated to my partner Manish and all those people, who helped me while doing this project.

TABLE OF CONTENTS

Abstract

The internship at Medulance exposed me to the startup culture and helped me understand the risks of Medical in the initial stages and the amount of work and dedication that is required to bring the product live. I gained valuable insights about how a product is prototyped and how the technical team works and how all the things are managed and developed upon by team members. I learnt about how to analyse & visualize the data from which we can make predictions. And learned how to clean & analyse the data, which comes under exploratory data analysis & finally visualize using libraries & various others tools. I am glad to have learned to develop in this framework through this Internship.

About the Company

Medulance is a Startup company based in Delhi and upgrading and digitalizing the traditional channels that were used to seek help. India's first, GPS based technology platform for fast and reliable first point medical attention.With an increasing emphasis on promoting independent living today, having access to the nearest ambulance to you can provide much needed peace of mind in a worst case scenario.

Our technology platforms were built to connect the customers and the ambulance drivers directly with each other in the time of need. We strive to provide high-quality, clean ambulances at competitive prices within minutes, assisting the every day superheroes to reach the patients faster with the help of an easy and simple format. We believe that ambulances should be able to help you when it matter

Our mission is to inspire breakthroughs in the way India looks at ambulances and first point medical attentions, and to touch lives. Our values serve as a compass for our actions and describe how we behave in the world.

SERVICES

a. Data Analytics

Data analysis solution which work on principle of 'Identification of the problem', 'Designing Data Requirement', 'Pre-Processing Data', 'Performing Analytic over data and 'Visualization of data'.

b. To make ambulances available for 24hrs

Our technology platforms were built to connect the customers and the ambulance drivers directly with each other in the time of need. We strive to provide high-quality, clean ambulances at competitive prices within minutes, assisting the every day superheroes to reach the patients faster with the help of an easy and simple format.

c. Medical Healthcare

India's Leading comprehensive emergency response service provider

c. Managed IT services

All domains including application development

- Time and Material based contracts

- SLA based contracts

d. Automation Services

Development of automation solution, which includes monitoring of assets, navigation, quality inspection and many more

Technology Stack

Python

This programming language is used for making chatbot.

Flutter

Develop applications for Android, iOS, Linux, Mac, Windows, Google Fuchsia and the web from a single codebase. The first version of Flutter was known as codename "Sky" and ran on the Android operating system.
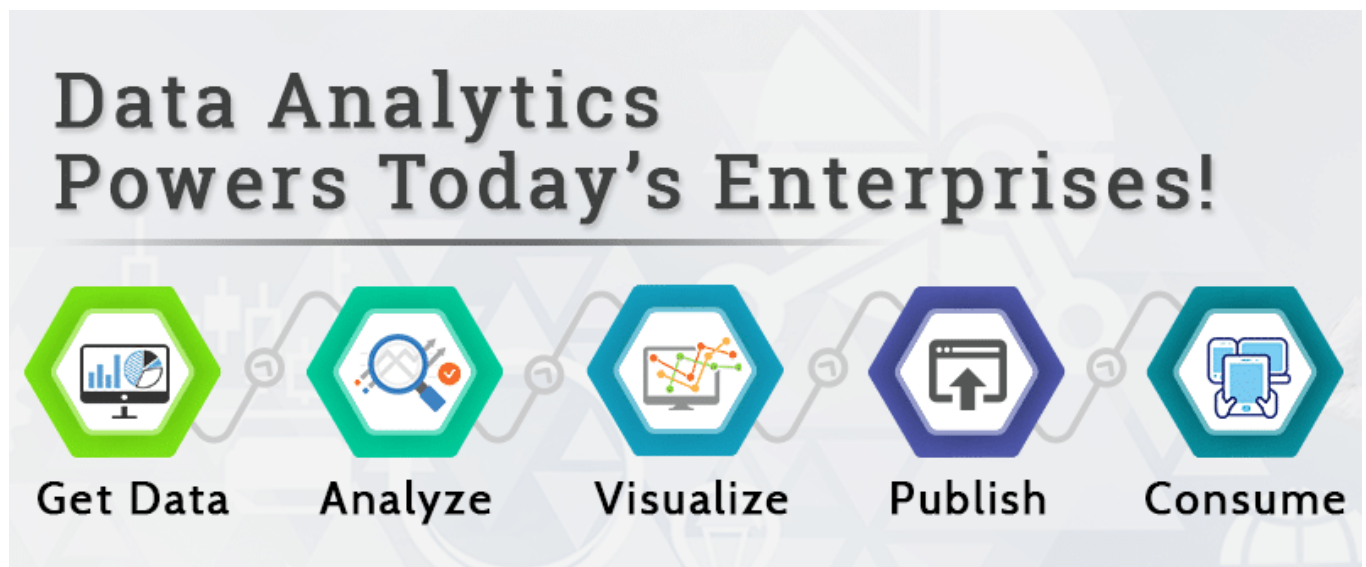
Data Analytics

Get to know the regions with most disease people in nearby area by using some analysis & making predictons & supporting decision-making

Data Analytics Domain

Data Analytics refers to the set of quantitative and qualitative approaches for deriving valuable insights from data. It involves many processes that include extracting data and categorizing it in order to derive various patterns, relations, connections, and other such valuable insights from it.

Today, almost every organization has morphed itself into a data-driven organization, and this means that they are deploying an approach to collect more data that is related to their customers, markets, and business processes. This data is then categorized, stored, and analyzed to make sense out of it and derive valuable insights from it.

The larger the size of the data the bigger the problem. So, big data may be defined as the data the size of which itself poses the problem and which needs newer ways of handling it. So, the analysis of data at high volume, velocity, and variety means that the traditional methods of working with data would not apply here.

<u>**Knowledge Acquired from Training**</u>

Data Analysis & Visualization

This project deals with covid-19 data analysis on datasets & analyze & predict the patterns for up-coming future.
And how this analysis can be helped in how much will it spread & which are most dangerous locations, called as hotspot

To help them in project, they told me to learn some basic tools & technologies, which will be further needed while performing exploratory data analysis
& visualize the graphs & charts.

Process include :
- Data Collection
- (EDA)Exploratory data analysis(Feature engineering,Feature Selection,Preprocessing,data cleaning,handling missing values in dataste)
- Model building(ML/DL) & selction suitable algorithm(supervised,un-supervised,reinforement) model for training,testing,Accuracy
- Deployment using AWS & Heroku, everyone can have access & be public
- PPT making, how your prediction can help company from other competitor & benefit & rise in economy

Feature engineering(Pre-processing) & feature selection, for data cleaning, this is step/process covers 80% of major work & it is most important part of timeline.

Data cleaning(handling missing vlaues) can be done in 3 ways:
- By taking mean,median,mode of data
- By imputation
- By replacing most repeated values

This can be done throughly with good understanding & analysing the data, what it is saying.
This process require, numpy,pandas,matplotlib,seaborn & some statistics techniques/topics needed & pick the best approach.

I applied all ML algorithmns(SVM,Naive bayes,Linear,Logistic,kNN,Kmeans,Random forest) to measure best accuracy,i.e., which one is providing more accurate & perfect predicton, we choose that algorithm
Performed some trainig & testing on models & validate the results.

After EDA & model building, I need to visualize the data, what is it saying?
What will be covi-19 predictions? how much will it spread?

All these questions & thoughts needs to be taken care while visualizing the graphs & charts, which is output of analysis
So, prepared a list of questions & answers to express my understanding & thoughts of visualizations I got & make them understand in best possible way.

They told to make a PPT & include all your process, steps, & timeline & your views, thought.
This was require because they want to test my communication skills, how well I can express myself to my colleagues & company benefit.

I made PPT & tried my best to understand them my all the cycle of data analysis & visualizations of project.
And also how my accurate predictions can help in determining how much more covid-19 spread in upcoming months.

Work Details

Date	Activity
15-Apr-2020	Joined the company & uderstood the environment & had interaction with collegues. All this was virtually as due to covid-19 pandemic
16-Apr-2020	They told to revise all the basic concepts of data analysis & visulaizaton frameworks,technologies,tools & libraries
17-Apr-2020	They told me to do some basic program in python given by them. They wanted to tets my skills.
18-Apr-2020	They explained me about their project which is on covid-19 data analysis on datasets & analyze & predict the patterns for up-coming future. And how this analysis can be helped in how much will it spread & which are most dangerous locations, called as hotspot
19-Apr-2020	To help them in project, they told me to learn some basic tools & technologies,which will be further needed while performing exploratory data analysis & visualize the graphs & charts.
20-Apr-2020	Then I started learning workflow or timeline of Data Analysis,i.e, Data Collection (EDA)Exploratory data analysis(Feature engineering,Feature Selection,Preprocessing,data cleaning,handling missing values in dataste) Model building(ML/DL) & selction suitable algorithm(supervised,un-supervised,reinforement) model for training,testing,Accuracy Deployment using AWS & Heroku, everyone can have access & be public PPT making, how your prediction can help company from other competitor & benefit & rise in economy
21-Apr-2020	This long time-line took some time to learn & apply in real project. At start it look very hard & difficult to understand. I learned little bit of some timeline.
22-Apr-2020	They fianlly assigned me project as i ahve gained some knowledge & understanding of workflow of any project cycle.

23-Apr-2020	At starting, I was not assigned any dataset, So i need to search on internet. For that data collection play role. The data I got in documented(JSON/No-SQL-MongoDB) form, means it was in non-structured form
24-Apr-2020	I scraped all the data requied from document using scipy tool/libraris(beautifulsoap,requests,urllib)
25-Apr-2020	After scraping, I need to clean my data & make it in proepr form,means row-column table form(SQL form). For that i need to know how to read & write the non-structured data into more proper form.
26-Apr-2020	After trying many attempts, I finally converted into excel sheet in much more better & proper form. Now all rows & colums have data inside.
27-Apr-2020	But,the data was not good enough to make good prediction from analysis as there are so many noisy data, missing values,ect... I have to clean the data & make it in more proper form
28-Apr-2020	Here comes,Feature engineering(Pre-processing) & feature selection, for data cleaning. I some libries & removed all the erros,bug,free spaces, missing values,etc..
29-Apr-2020	This step/process covers 80% of major work & it is most important part of timeline. If mistake happens, your predicton can go wrong & company may loose benefit.
30-Apr-2020	This can be done throughly with good understanding & analysing the data, what it is saying. This process require, numpy,pandas,matplotlib,seaborn & some statistics techniques/topics needed & pick the best approach.

Date	Notes
01-May-2020	Data cleaning(handling missing vlaues) can be done in 3 ways: By taking mean,median,mode of data By imputation By replacing most repeated values
02-May-2020	Till,here I have learned how to collect data(structured or non-structured) form & perform Exploratory data analysis(EDA) i.e., Pre-processing & feature selection, for data cleaning.
03-May-2020	Now, I need build model,can be ML model or DL model, which can be helpful in trainig,testing & measuring accuracy of model All the data need to be feed to model to predict better accuracy.
04-May-2020	I applied all ML algorithmns(SVM,Naive bayes,Linear,Logistic,kNN,Kmeans,Random forest) to measure best accuracy,i.e., which one is providing more accurate & perfect predicton, we choose that algorithm Performed some trainig & testing on models & validate the results.
05-May-2020	After comparing all models accuracy, ranom forest gives best prediction. So I selected that model
06-May-2020	I got to know to some insights of how model works & what are the things required to better results. I faced some bugs & errors but collegues helped me to fix bugs.
07-May-2020	Now, model prediction was going right & was giving better results for accurate prediction. But I tested the code again, if there is some bugs which are efficency of model.
08-May-2020	Again, I performed EDA & trained the model & tested against trained mode. And yes there was some percenatge of increment in accuracy. Which shows that, there was some bugs & data was still needed to more cleand than previous

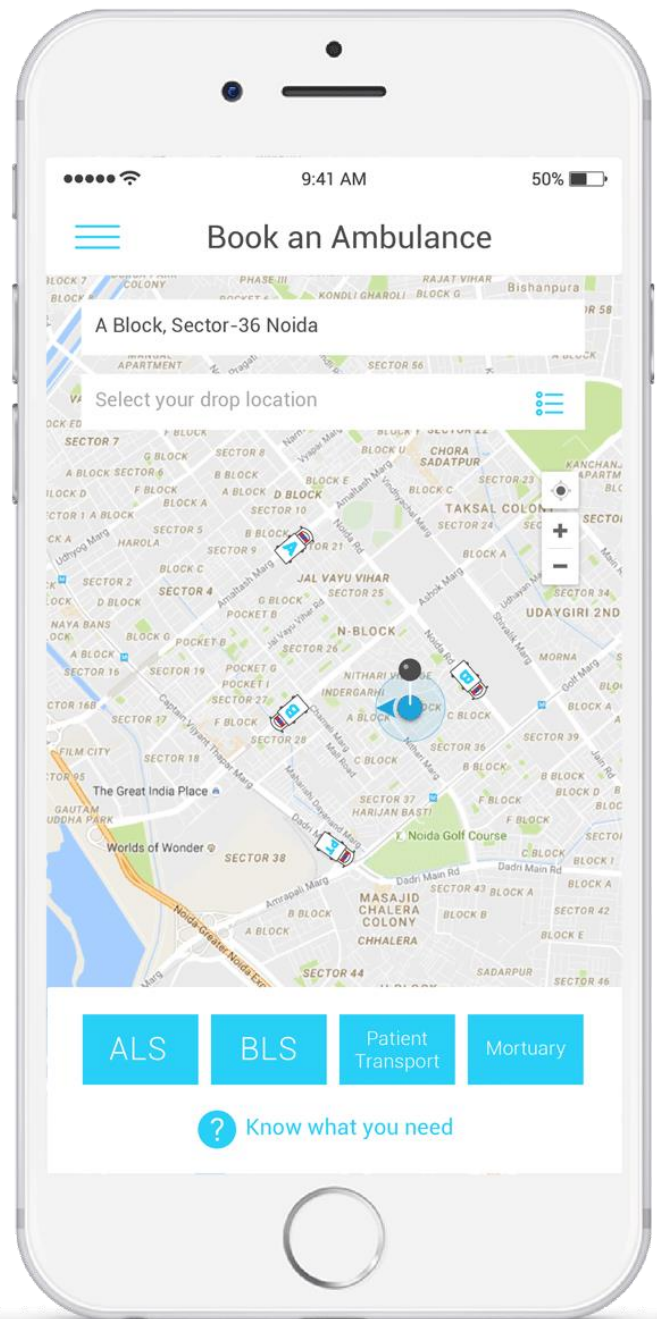|	
09-May-2020	This is all due to some colums & rows of data left uncleaned & needed more analysis of data. So, I got more accuracy than previous result. Everyone happy with my work by improving more better & accrate data for prediction.
10-May-2020	After EDA & model building, I need to visualize the data, what is it saying? What will be covi-19 predictiosn? how much will it spread?
11-May-2020	All these questions & thoughts needs to be taken care while visualizing the graphs & charts,which is output of analysis So, prepared a list of questions & anaswers to express my understanding & thoughts of visualizations I got & make them understand in best possible way.
12-May-2020	They told to make a PPT & include all your process,steps, & timeline & your views,thought. This was require because they want to test my communication skills, how well I can express myself to my colleagues & company benefit. I made PPT & tried my best to understand them my all the cycle of data analysis & viuslizations of project. And also how my accurate predictions can help in determing how much more covid-19 spread in up-coming months. Till, here I have performed well & gained some idea or workflow of project. And there was imporovement in project. It took nearly 20 days to master 80% of work.This is all due to my colleagues who guided & helped me a lot in this journey

Application

Features of Medical HealthCare Domain

- Safe: Avoiding harm to patients from the care that is intended to help them.

- Effective: Providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and misuse, respectively).

- Patient-centered: Providing care that is respectful of and responsive to individual patient preferences, need

- GPS based technology platform for fast and reliable first point medical attention.

- With an increasing emphasis on promoting independent living today, having access to the nearest ambulance to you can provide much needed peace of mind in a worst case scenario.

- You can book from anywhere

- Its flexible & user friendly interface

- One touch access to medical emergency services.

- • Seamless communication with ambulance drivers

Market Adoption

- You can book from anywhere

- Its flexible & user friendly interface

- Previous years, our application helped a lot & been used my many peoples

- Take immediate action & no delay in transporting

- GPS based technology platform for fast and reliable first point medical attention.

- With an increasing emphasis on promoting independent living today, having access to the nearest ambulance to you can provide much needed peace of mind in a worst case scenario.

- One touch access to medical emergency services.

- Seamless communication with ambulance drivers

- Keep your dear ones close with emergency contact option.

- Convenient mechanism, with online and cash payment mode.

Conclusion

The given tasks were successfully implemented in the internship period. It was overall a good learning experience. I learned new tools, frameworks and experience of working with Medulance team.

During this period, I worked in various areas like python and data analytics successfully met the objectives that were set at the beginning of the project.

I am thankful to Medulance for providing me this opportunity to develop and apply my skills. I am grateful to my mentors for providing me the necessary training and guidance. It was a very valuable training, in which I learned this popular skillset which is high in demand in robotics. The experience and skills gained will surely be of great benefit in my career.