

# Sarthak Taneja

Gurgaon, Haryana, India 📌

+91 9717988155  
sarthaktaneja1996@gmail.com

[Know more about me](#) 🐝

## Data Analyst

### About

Gathering, analysing and the art of storytelling about data is what pushes me to work hard and grow my skills and make data backed decisions everyday.

### Experience

#### Data Analyst, Ernst and Young India

Gurgaon – Aug 2019 to Present

Part of the Innovation team building machine learning and artificial intelligence assets. Collaborating with financial and government institutions on projects related to data modelling, data visualisations, Natural Language Processing (NLP), document digitisation using OCR and ICR tools etc.

#### Android Developer, Paytm

Noida – Jan 2018 to July 2018

Android and web development for Paytm and Paytm Mall applications.

#### Android Developer, OKS

New Delhi – Jun 2017- July 2017

Created an Android application for simulations of science concepts for high school students.

### Education

#### Delhi Public School

Gurgaon – Matriculation, 2012

#### Delhi Public School

Gurgaon – Higher Secondary, 2014

#### Symbiosis International University

Pune – B.Tech CS&E, 2015 to 2019

## Skills

Python	<div><div></div></div>	Android	<div><div></div></div>
HTML	<div><div></div></div>	CSS	<div><div></div></div>
Azure	<div><div></div></div>	SQL	<div><div></div></div>
Power BI	<div><div></div></div>	Tableau	<div><div></div></div>

## Academic Projects

### Foody

A database in SQL that lets the user find a restaurant based on the choice of cuisine, price, ratings and location. This project taught me how to create tables in SQL and showed how various SQL queries are written.

### Malware Detection And Analysis of DOC Files

A machine learning based system to detect malware in .doc extension files using decision trees and SVM by static analysis. Also studied about various malwares and techniques to remove the obfuscated parts of the malicious file.

### Android: All you need to know

#### Research

Investigation on how the android OS has improved over the years, what changes have been made in its architecture & structure and the new versions introduced.

### Blockchain for Electronic Health Records (EHR)

#### Research

Instead of relying on a designated intermediary for information exchange, such as a state-designated HIE (Health Information Exchange) or a private network established between local hospitals, the decentralised nature of the blockchain would allow any approved participants to join an exchange community, without the need to build data exchange pipes between certain organisations.

Presented as a paper titled 'Blockchain for Electronic Health Records (EHR)' at IC-ISCN 2019 hosted by the Springer Conference in Mumbai.

## **Professional Projects**

### **Propensity to not pay mobile bills**

Preventive detection tool for non-payment of mobile bills, determine the probability of non payment by customers, on invoices before they are due. Classifying a customer into pre defined categories based on their historical information. Python and its libraries were used for Modelling, Data visualisation and feature engineering. The client shared GBs of customer information in flat files which was a major challenge so we used Microsoft Azure Data Factory for faster and easier table creation in SQL Server. Also there were around 700 variables which were provided to us by the client. The XGBoost algorithm gave us the best results which yielded an AUC score of 0.84.

### **COVID Resource Planning Tool**

Need of the hour when the pandemic struck was to have resources (beds, cleaning staff, PPEs etc.) allocated to hospitals/ districts which needed them the most. The COVID resource allocation tool was used to predict the number of covid cases in the future using the epidemiological models SIR and SEIR. The development was done using Python and was deployed on Azure. The functionality of the tool was to either change the parameters like population, doubling time and other factors in real time or bulk upload the data points using excel. The final predictions were fed into a resource planning tool Anaplan. Also, a dashboard was build on Power BI to analyse the situation in that particular hospital/ district. The biggest challenge was gathering latest COVID data from credible sources.

### **Executive dashboard for a Leading Credit Card Company**

The executive management at a leading credit company wanted to view their customer care specialists' performance metrics which was earlier shared as a large excel table which was inefficient to view and analyse. The biggest challenge was collating data from various teams and systems. Also, row level security was little tricky as there were multiple tiers of the management authorities. Finally, after ironing out theses challenges we were able to collate the data in a csv file which was the input to Tableau.

## **Interests**

Travel, Books, Movies, Photography, Music, Sports