Personal Statement

I've done several projects in data science which can be found in my Github repository. One of the most recent was **Accessing and analysing UK covid-19 Data from UK Open Data API (Plus Global COVID-19)** accessible at the following links: https://tinyurl.com/2p8ykby9

I embarked on this project when I was applying for a data scientist position at the UK Cabinet Office. The potential candidate was to join the Covid-19 Task Force. I did a Google search for covid-19 data for UK and the search took me to UK OpenData API for developers. I followed the instructions on the site to harvest UK Covid-19 data and performed a quick analysis and uploaded my results to Github before applying for the position. I presented the results of analysis using graphs which are very self-explanatory for both technical and non-technical folks. As part of the story telling, I also provided some comments and thoughts which can be helpful to the stakeholders and decision makers, with a recommendation to the UK government to halt all mandatory testing! Tests should only be conducted on patients as part of diagnosis and treatment. This is because with low prevalence of the disease most positive test results are false positives. This is due to irreducible error in the test.

If statistics were a pair of shoes, I would claim that I know where it pinches!

In most of the courses I've taken on Udemy and in all the books I've read, it's claimed that one needs not so much of statistics to be able to follow up. However, after doing a number of projects I realized that one needs more statistics than anything else to be successful in this domain. I stepped back to make sure I understood statistics beyond the level of a classroom!

I read and re-coded into Python a 600+ page book on statistics which used SPSS in its examples. I decided to use Python to solve all the examples and end of chapter exercises! I now feel comfortable applying statistics in a practical sense. I applied my updated knowledge on a real-world dataset from The National Health and Nutrition Examination Surveys (NHANES). I conducted a series of statistical hypothesis tests and the result can be seen at:

https://tinyurl.com/ychbrb5p and https://tinyurl.com/uc4ckhhw

In some of the tests the results were statistically significant but I used the Effect Size to overturn the outcome and fail to reject the null hypothesis.

I used to teach Math and Physics in a secondary school and my former students continue to make me to feel like I was one of their best teachers! This is because I have the patience to explain any difficult concept to the level of common sense! I am therefore a suitable candidate to mentor junior colleagues.

I was selected by UK DWP for an interview. The job adverts indicated that the department deals with fraud detection. Therefore, I downloaded and analysed a fraud detection dataset. I used statistical hypothesis test to select the good predictors of the target variable. Then I trained a logistic regression model which could out-perform the manual, inefficient and unsustainable process of detecting fraud as follows:

- reduce significantly the amount of backlog (incoming and unfinished work),
- reduce drastically the number of staff needed to do the job,
- shortened the processing time significantly,
- more cases of fraudulent transactions would be tracked down in a given amount of data processed - more than 40% increase in efficiency as shown by the Cumulative Accuracy Profile (CAP) curve!

I was given a similar Case Study at interview! I needed no further preparation for it was a direct transfer of knowledge and I hope my interviewers were impressed!

The Github repo is **Fraud Detection Model** accessible at https://tinyurl.com/37m6d5ap

From the look of things, the model could be improved by applying PCA to generate a single feature for the logistic regression model.

I beat a Kaggle Competition!

An instructor in a course on Udemy gave a challenge that we should see whether we could obtain a better RMSLE for a past Kaggle Competition ***"Predicting the sale price of bulldozers sold at auctions"*** using Random Forest Regression. I was the only student who reported a score better than that which won the competition. This is how the instructor reacted to the results *"woahhh! What an epic function! I love the handling of numerical/categorical/missing columns sequentially as well as the use of ordered feature names!"* This was in regards to the approach I took to engineer the features! The notebook can be found on my Github repo at http://tinyurl.com/zvp3npha.

As part of an interview for a data scientist at 10 Downing Street I had to undertake a data science technical exercise. I analysed data on **National Tutoring Programme (NTP)** as practice exercise and on the date of the interview the assignment was analysing data on **Accident and Emergency (A&E) Performance**. I concluded the exercise with recommendations to UK Department for Education of 10 Local Authorities where NTP should be intensified. For the A&E performance target I conducted statistical hypothesis tests and provided a response to UK Secretary of State for Health with a conclusion that according to the available data the 76% performance target in A&E was a far-fetched goal. Github repo at https://rb.gy/9v47r8.

Image classification using German Traffic Signs Dataset (combat overfitting using data augmentation)
Github repo: https://tinyurl.com/4nbjjzrv

My goal was to beat the best reported score but unfortunately, my laptop does not have sufficient memory and processing power real estate. A good laptop with GPU would cost me about £2500. Immediately I'm able to afford one I would take the ***Practical Deep Learning*** course (https://course.fast.ai/) by Jeremy Howard and apply the acquired transfer learning skills to improve the score.

Using complex SQL queries to answer specific business questions. Notably using multiple named subqueries, views to extract data from a database to address specific problems.
Github repo at https://tinyurl.com/4xd6xryr.

As GHG auditors our job was to travel around the world, mostly to developing countries, to conduct validation and verification audits. The clients or project owners would enter into a contractual agreement with us and we had to agree on a timeline for deliverables, with penalties if any deadline was missed! So, in order not to miss a deadline and lose money we had to work and, at times, under very tight schedules. We used to run into troubles at times because the UN requirements were constantly being revised and before we could complete a project some requirements that were applicable at project start have changed and we had to revise some sections of our reports to be in line with the new requirements. We had to be agile! And since I was the technical review manager the burden was always on me as the last

person to review and release the reports. I am proud to say that my record at the UN as QM is green! My clients were never disappointed with me or my performance. I remember my boss asking me what the source of my motivation was. I told him results motivate me and I don't have any reason to under-perform!

Still in this role I had a medium size team of auditors to manage, mostly remotely. We had auditors in China, India, Indonesia, Brazil, Mexico and Europe. In this role part of my responsibilities was to monitor the ever-changing UN requirements, updating the team on a bi-weekly basis and also updating our QMS to reflect those changes. Barely weeks into my new role my boss noticed a change in attitude in the team and came asking what I had done to bring about the positive change. Even our clients were writing positive things about us he said. Then he asked me whether I was applying my skills as former teacher in my work? I told him that my personal background and the experience as a teacher was certainly an asset. I know how to appreciate and I know how to motivate someone to achieve even more. And most importantly I have cultivated the attitude to be patient, to listen to others' opinion and the humility to accept the fact that one can't be right all the time.

I also used to plan and conduct internal audits and address any issues found. Since our business unit was accredited by UN we had to absorb external audits by UN experts. I was the focal point for any raised corrective action requests which I had to respond to with proposals on how we intend to address them. This involved making some major decisions and communicating such to UN. We had a maximum of two rounds of communications to address the issues and if unsuccessful after the second round we would get a suspension! To avoid this, I had to make sure that the decisions I took and proposed to the UN were clear and would address the issues to the satisfaction of all parties. When I finally got the opportunity to be on top of the decision hierarchy, I use to triage the issues raised in the first round of communications. This was new to my boss who was fond of betting with me!