

Ioana Fiona PREOTEASA

Data Scientist

[in linkedin.com/in/ioanapr](https://www.linkedin.com/in/ioanapr) github.com/loana-P
+447845698765 @ ioanacpreoteasa@gmail.com
London

Data scientist experienced in working with data on fragile and conflict-afflicted contexts. Have developed full data science pipelines for projects ranging from regression modelling of migration flows to ensemble machine learning classifiers of human trafficking articles. Worked on a variety of pilots with HMG's Conflict, Stability and Security Fund (CSSF). I turn opaque, heterogeneous data into clear **narrative**, actionable **insights** and robust **models**. Able to **visualise** and expedite key information to stakeholders in both Python and R.

SKILLS

Programming languages	Python, R, SQL, git
Libraries	Pandas, Scikit-Learn, Pytorch, Keras, Numpy, NLTK, RegEx, statsmodels, imblearn
Visualisation	ggplot2, Matplotlib, Seaborn, Streamlit dashboarding, plotly
Databases	SQLite3, MongoDB
Cloud development	Amazon Web Services, Google Cloud Platform
Soft skills	Project management, public speaking, scrum-group management, pedagogy
Writing	Published in Towards Data Science and Analytics Vidhya
Other Software	Visual Studio Code, Jupyter Lab, RStudio, Google Collab

PROFESSIONAL EXPERIENCE

Current
Jan 2021

Consultant Data Analyst, FREELANCE - INTEGRITY GLOBAL, London, UK

- > Developed a forecasting model for Integrity's CSSF clients over several weeks. To enable **repeatability, transparency** and **further experimentation**, I refactored the prototype model code into a **reproducible analytical pipeline**. We were thus able to produce over 30 iterations of new models in three days, complete with visualisations, statistical analysis, model files and performance summary tables and add easily incorporate new data when it becomes available. As a result we were able to deliver a model with significantly lower RMSE than our MVP, and further validate our modelling.
- > Went above and beyond the client's requests: developed an additional **statistical model** on a different target dataset that I anticipated the client would be interested in predicting. This outperformed the best previous model and generated further insights into the problem area for the client. I used my **contextual knowledge** to preempt the models' assumptions and limitations without understating its value to the client.
- > Produced a Python-based programme with a **Guided-User-Interface(GUI)** that allows a user to extract information from pre-formatted MS Word documents and output usable, conditionally formatted tidy data files and pivot tables. Effectively cut a manual extraction process that normally took **several days to a matter of minutes**. Also wrote three separate pieces of handover documentation for various levels of technical proficiency, and planned and delivered a technical workshop for those taking our code on the client's side.
- > Successfully **championed the greater adoption of git and GitHub** within Integrity's CSSF and World Bank projects. Took the lead on the documentation and upskilling required for collaborating data analysts. Reduced lengthy and error-prone sharing of code that involved emailing files to an easily managed and transparent git workflow.
- > I took the initiative in synthesizing a literature review of the field of conflict forecasting into a five page document. I successfully translated complex information about modelling for a non-statistical audience using insights ranging from ongoing academic models of conflict to ideas of utilising hybrid human expert-machine systems.

ggplot2 time series modelling forecasting research report writing

Current May 2021	<p>Consultant Data Engineer, FREELANCE - INTEGRITY GLOBAL, London, UK</p> <ul style="list-style-type: none"> > Built a functioning ETL, QA and data merging pipeline. This produced weekly survey data intended for the World Bank. The QA script outputted several Excel files with multiple tabs, showing highlighted, commented cells so the monitoring team could understand and rectify / confirm the errors. The data merging script combined the datasets and converted them to a specific client-friendly format. I reduced several days of difficult manual QA that would sometimes take over half of the weekly reporting cycle to a few hours of plug-and-run automation. > Refactored and lead further development on aforementioned QA scripts so that they can now be recycled and easily adapted for future projects. > Automated the manual work of qualitative analysts who were spending 1-2 days sifting through thousands of rows of client data to identify references to the Taliban and related entities. My scripts allowed them to receive their curated data extract within an hour of us receiving the whole data from the data collection agency, and gave them more time to analyse the qualitative data in detail. <p>Python Pandas Excel Quality Assurance ETL regex GitHub</p>
Jan 2021 Sep 2020	<p>Data Mentor, DECODED, London, UK</p> <ul style="list-style-type: none"> > Mentored a cohort of data analyst apprentices in learning and using data science concepts and techniques to solve problems in their day to day work. <p>Python Pandas Numpy project management SQL Seaborn Matplotlib Plotly Zoom Google Hangouts</p>
Jun 2020 Feb 2020	<p>Data Science Coach, FLATIRON SCHOOL, London, UK</p> <ul style="list-style-type: none"> > Led multiple student cohorts to success in their final portfolio projects, on a range of topics from image recognition to recommendation systems. > Delivered remote lectures on NLP, neural networks and briefer tutorials on topics such as Support Vector Machines, Principal Component Analysis. <p>Python Pandas Numpy Keras git Github project management SQL Seaborn Matplotlib GCP Zoom</p>
Jul 2019 Sep 2017	<p>Science Teacher Teach First Participant, DUKE'S ALDRIDGE ACADEMY, London, UK</p> <ul style="list-style-type: none"> > Developed pedagogical and leadership skills through two years of working full time as a teacher and studying towards a vocational qualification (PGDE - Postgraduate Diploma in Education). > Designed, delivered and evaluated research and evidence-informed interventions in my classrooms. <p>pedagogy academic research intervention design presentations public speaking leadership</p>

VOLUNTEERING PROJECTS

Feb 2021 Nov 2019	<p>DataDive Volunteer, DATAKINDUK, London, UK</p> <ul style="list-style-type: none"> > Developed a web-scraping and data modelling pipeline for an anti-trafficking organisation (Stop the Traffik), as part of an 12-month data ambassador project. > Participated in three separate DataDives during which I was briefed on a charity's background and data-related problem, given access to their data and tasked to deliver a product in under 36 hours. > Feature engineered client data and then performed k-means clustering on those features. Assesed clusters using elbow method and Silhouette Analysis. Interpreted 5-dimensional clusters using a range of visualisation techniques and interpreted results for client (St John's Ambulance). Wrote a very brief report on insights and recommendations for future research, under severe time constraints. > Built and optimized classification models for an educational charity partner, The Brilliant Club. Reported on the final model's sensitivity (True Negative Rate) and visualized what the most predictive features of their data were for solving their problem. > Using topic modelling I produced insights for a Scottish 3rd sector company (FirstPort), based on their survey data. Explained the interpretation, limitations of the model to the stakeholders and provided them with an interactive topic model dashboard (created via pyLDAviz). <p>Decision Trees Random Forest ScikitLearn Seaborn Matplotlib Clustering Unsupervised Learning k-Means Latent Dirichlet Allocation Zoom Slack web-scraping pipeline development</p>
----------------------	--

EDUCATION

2020	Flatiron School - Data Science 15 Week Course
2019	Postgraduate Diploma in Education, University College London's Institute of Education
2016	Bachelors of Science in Theoretical Physics , University of York
2013	Lancaster Royal Grammar School