

Junior Applied Data Scientist

Salary

£30-40k depending on experience.

Up to 0.1% equity depending on experience & salary, equity/salary exchange negotiable.

Applications reviewed as they come - so apply quick

What you'll be doing

Building MRP models and analysing outputs

- Preparing raw survey data for modelling.
- Carrying out bespoke exploratory analysis.
- Finding and collecting relevant data from external sources.
- Running and managing our proprietary MRP pipeline.
- Tuning and refining MRP models to optimise their accuracy.
- Validating and analysis of model predictions.
- Visualising results for our clients.

Data processing & ongoing automation

- Analysing / validating / filtering of raw data to remove poor quality respondents.
- Cleaning of raw data into formats / structures suitable for downstream modelling.
- Building prototypes that automate (a) the processing of cleaned data into weighted data tables and (b) data quality checks.

Customer project delivery

- Independently manage project workload and deliver client results to tight deadlines.
- Build Powerpoint presentations.
- Make graphs and visualizations.
- Respond to client queries around survey results and modelling methodologies.
- Provide proofreading and accuracy checks for customer surveys, abstracts, presentations, reports, and manuscripts.
- Assist with the development of survey questions.

Data science & analysis

- Work on projects such as: customer segmentation, profiling, clustering and customer journey analysis, regression.
- Combining surveys together and analysing for improvements to the product.
- Applying your initiative to improve existing processes - get creative! We encourage you to automate aspects of your role using any clever ideas you might have.

Managing our internal statistics repository

- Finding and extracting data from public data sources / APIs to be used for MRP modelling, to be added to our internal catalogue of national statistics, for countries around the globe.
- Writing and incorporating new ETL scripts into our statistics-management repository to expand the catalogue for these new datasets.
- Using and developing novel techniques to ensure internal consistency within the catalogue, and to infer statistics between different geographies.
- Maintaining and improving the CI/CD pipeline we have in place to manage and update our statistics.

Input into product and engineering roadmap

- Distilling your own experience using our stack into feedback to help us build / improve features, in particular around data quality and data analysis.
- Communicating problems with infrastructure to the engineering team.

You should have - technical skills

- 1+ years of relevant experience as an analyst / data scientist or relevant quantitative degree (preferably MSc). Enthusiastic graduates with a strong work ethic are also encouraged to apply.
- Experience with the full data analysis lifecycle in Python or R - ingestion / cleaning / transformation / validation / etc.
- You are comfortable using one of the standard munging libraries (pandas / dplyr / data.table / ...) and have experience using at least one of these tools in a professional environment.
- You are comfortable scripting for data cleaning / manipulation / analysis, and generate code that is readable, reproducible and reliable.

Nice to have:

- Experience with SQL.
- Experience in data visualisation.
- Experience in communicating technical materials (e.g. machine learning models) to non-technical audiences, through presentations, blog posts, or written reports.
- Experience with Google Cloud Platform / AWS / Microsoft Azure.

- Familiarity with survey methodologies.

You should have - mindset and non-technical skills

- Desire to work in a startup environment and an understanding of what this entails: you are unfazed by uncertainty; you're happy to work across different parts of the business as and when required.
- You are willing to work hard when there are important deadlines to meet (especially when working with customers).
- You are keen to learn new technologies, domains, etc. - we're here to support you in all your learning efforts!
- Strong attention to detail.
- Critical thinking - you naturally ask questions about data, think about how to make data as easy as possible for customers, etc.
- Proactive attitude - if there's anything you want changed, you can make it happen
- You are able to communicate clearly with a variety of people - both within the team and with clients.

Nice to have:

- Interested in survey data and enjoy playing with raw data.
- Interested in UK and global politics (or current affairs).
- Experience in market research.
- Appreciation that data analysis is hard and takes time.

About the company

We use machine learning to understand public opinion better than traditional polling. Our insights empower campaigners, governments, brands and CEOs with cutting-edge analytics on consumer opinion, helping them make better decisions. Our work has been featured in various media outlets: for example, [Channel 4 News](#), [Wired Magazine](#), the [Financial Times](#), and [The Economist](#), to name a few.

Winner of a CogX award for best AI product, we have raised over £1.4m in funding from the founders of Zoopla, LoveFiLM, Alchemy Partners.

About our technology

We process thousands of survey questions through our data pipeline, to infer public opinions to a synthetic proxy of the national population. The majority of the data science pipeline is written in Python. Downstream we use Apache Spark, Apache Beam and Google Dataflow to transform the massive amounts of data from our model predictions to make it accessible to our analysts and clients.

Our customers explore their results through our slick web application. The majority of this stack is written in TypeScript, React, GraphQL, whereas our backend is mainly written in Scala using different types of databases including but not limited to Postgres, Redis and BigQuery. We have adopted GitOps, using Git as our single source of truth, to manage all our applications, infrastructure, access rights, and deployments. All our infrastructure is managed in Terraform and all our services are running on Kubernetes. This allows us to recreate our whole system within 60 minutes. Our architecture is based around the concepts of microservices using Apache Kafka as event-store and event bus as well as Istio as Service Mesh and API gateway.

We believe in modern software engineering principles, including DevOps, SRE and agile software delivery practices. This ensures that our customers are happy and the continual growth of our engineering team.

Here are some of the projects and some of our big engineering challenges we are working on right now.

- Process survey responses in real-time to analyse and transform the data for downstream services (billing, monitoring, real-time data visualisation), automatically making it available to our customer analytics application, and also feeding our machine learning pipeline.
- Keep our data consistent across different microservices and data stores using Change Data Captures and Apache Kafka.
- Provide the highest level of data quality to our customers using machine learning and real-time data analytics.
- Optimise our respondent and market research web application for speed, usability and user experience, so we meet the ever growing expectations of our customers and users.
- Automating the feedback loop between deployment pipeline, software and infrastructure. So our team will know of software defects before they reach production and before our customers notice.

Our mission

Our mission is to build an intelligent system for understanding consumers that powers better decisions by the world's most important organisations. We provide our customers with more accurate and granular predictions in a shorter turnaround. Our customers span a wide range of diverse industries - from retail, to politics.

Benefits

- Flexible equity/salary exchange negotiable
- Bright plant-filled office in Old Street
- Christmas and summer parties (pandemic allowing)
- Working with a young, driven, friendly team