Runsang Yu

1011 Washington Blvd, Stamford, Connecticut, United States, 06901
Website: https://yurunsang.github.io/
Mobile: +1-2037081241 Email: yurunsang@icloud.com

PROFILE

Data Science Specialist with over eight years of experience working across diverse industries including FMCG, banking, and property consultancy. Armed with a robust tech stack encompassing Python, PySpark, SQL, and leading BI tools, my forte lies in leveraging data to drive business decisions through predictive modelling, business intelligence, and ETL process optimization. With a hands-on approach to data science coupled with a strategic business mindset, I've consistently spearheaded global projects, championed innovation, and crafted consumer-centric solutions that resonate and deliver.

WORKING EXPERIENCE

Philip Morris International

Manager Audience and Segmentation (full-time) — *Mar 2024 – Present*
Stamford, CT, United States

I am part of the US Marketing Operation team, leading audience segmentation, CDP strategy and data activation to support the launch and expansion of IQOS in the US market. I own the consumer data platform in Treasure Data, integrating data across CRM, eCommerce, paid media, customer care, retail and face-to-face channels to enable consistent, targeted consumer engagement. I design predictive and rules-based audiences (including life-stage segmentation, lead-scoring propensity models, zipcode-level "Treasure Map" and ZYN life-stage/competitive segmentation) that guide targeting, territory planning and indirect retailer expansion. Through close collaboration with commercial and consumer experience teams, I embed data-driven planning, measurement and test-and-learn into go-to-market activities to support sustainable growth.

Manager Segmentation and Modelling (full-time) — *Dec 2022 – Mar 2024*
London, United Kingdom

I work in the smoke-free product consumer experience team of a global function, I specialize in architecting predictive models, orchestrating strategic initiatives, and managing high-impact projects. With a strong focus on enhancing the consumer journey and experience, I leverage data-centric tools to drive insightful business decisions and operational efficiency. My experiences range from fostering global collaboration and mentoring interns, to revolutionizing operational processes and championing consumer-centric solutions.

1. Architecting Predictive Models and Driving Strategic Initiatives:

- Consumer Personalisation: Utilized Treasure Data (CDP) as a pivotal environment for data storage and established ETL workflows to streamline daily model scoring processes. With Python at the core of model building, I crafted dynamic consumer segmentation strategies to drive personalised engagement and increasing acquisition and retention.
- Predictive Analytics: Leveraging the capabilities of Treasure Data, AWS, and PySpark, I constructed and deployed predictive models for over 20 markets with over millions of consumers. These models, ranging from referral to churn and flavour advisory, were instrumental in driving strategic outcomes. For instance, the referral model alone resulted in a 3% uplift in acquisition, showcasing the tangible impact of data-driven decision-making.

2. Architecting MLOps & Advancing Feature Store Framework:

- Revolutionized the modelling process by designing a feature store framework within the Treasure Data stack. This framework encompassed 500+ reusable features around a comprehensive consumer 360 view, empowered with time-travel capabilities, and seamlessly integrated exploratory data analysis, model construction, and monitoring. This transformative approach eradicated inefficiencies stemming from redundant features and outdated practices.
- Introduced a robust Machine Learning Operations (MLOps) framework, integrating Git for version control and fostering a collaborative and efficient environment for data analysts and data scientists. This strategic shift significantly enhanced the data team's productivity and effectiveness.
- Presented the scalable ML framework at the Treasure Data 2024 Conference, showcasing its impact on streamlining model development and deployment across markets. (Video link here: https://youtu.be/jFfLJamxIgU?si=3rfcIkB90sFm43I3)

3. Spearheading Global Collaboration and Fostering Innovation:

- As a pivotal figure in the global function, I actively collaborated with market CRM leaders and market data scientists to educate them with the model development in local markets and regions and implementation, ensuring the models' strategic utilization.
- Partnered with global acquisition and retention teams to design customer journeys that effectively leveraged model outputs, thus enabling potent customer activation.
- Introduced forward-thinking initiatives such as the "Open Innovation," encouraging collaborative contributions from talented market professionals to enrich our global segmentation portfolio.

RBS International, NatWest Group

Data Scientist (full-time) *Nov 2021 – Nov 2022*

London, United Kingdom

Iteratively build and prototype data analysis pipelines to provide insights that will ultimately lead to production deployment via an agreed implementation approach. Seek opportunities to identify new methods, tools, techniques, and opportunities to deliver business value via cost reduction, income generation, or improved customer experience through the application of data science.

- New Branches Queueing Tool Simulation
  Developed a queueing tool in a web interface using Python to simulate the branches traffic under different scenarios to help stakeholders make decisions on closing or remaining the branch. Several parametres including footfall, different service points usage, branch max capacity etc could be selected by users to determine the optimal branch operation.
- Business Through Personal Accounts Identification
  Built a clustering model based on time series features to identify accounts that are making business transactions through personal accounts. The result was shown in an interactive dashboard to suggest past transaction details about these suspicious personal accounts to support decision making.
- Customer Mortgage Retention Analytics
  Built a classification model for mortgage customer behaviours when rolling onto standard variable rate products, predicted customer's likelihood to churn in the next three months, understood the drivers behind the model, and interpreted the results to senior stakeholders to suggest potential ways to retain the customers.

CACI Ltd

Data Scientist (full-time) *Sep 2018 – Sep 2021*

London, United Kingdom

1. **Date Modelling Projects:**
   - User Home/Work Locations Detection:
     Take lead on GPS data Projects on the technical side: extracting and processing data using Google Big Query, SQL Server, and Alteryx. Presenting data in Qlik dashboard including designing and updating the data models. Design streamline workflows for efficiency and help answer bespoke questions in terms of different client demands.
     - Developed people's favourite places using GPS location data and POI data to support leasing strategy and modelled out trip purposes to generate more insights for clients.
     - Modelled out thousands of users' home/work locations using DBscan based on millions of GPS location data. Combined the result with company's consumer segmentation products, to label each user, helping clients have a better understand of their customers' profiles in the assets.
   - Social Distancing model:
     Developed an Agent-based Model in NetLogo to simulate how clients can ensure their staff to return office safely under the social distance policy.
   - Sales Daily Forecasting Model/Customer Repurchase Likelihood Model:
     Developed machine learning models on Azure Databricks to forecast the daily sales for one of the biggest e-commerce companies and to predict the repurchase likelihood of new customers to help clients on marketing decisioning and boost the retention rate.
   - Consumer Segmentation Model:
     Used over ten thousand of questionnaire data to build a consumer segmentation model in the UK with a combination of unsupervised learning methods. The final results perfectly fit the UK market and are well received by UK clients.
   - Market Gap Model:
     Conducted machine learning modelling of the relationship between mobile flow data and bank data to predict financial transactions for the market gap analysis, achieving a high accuracy performance in terms of real market.

2. **ETL Process Design:**
   - Based on the context of building batch Tableau dashboards for over 40 shopping malls owned by the client, I led the team to design the whole ETL process in Alteryx and SQL resulting in a huge amount of time reduction (from a month for three people to one day for one person), and much higher accuracy.
   - Built the ETL process automation for one of the company's biggest benchmark products for all the UK shopping malls. Applied cleansing and processing routines, designed webpages to allow end-users to interact with different processing choices in Python, SQL and fed all the results into a Qlik dashboard. This hugely improved the speed of importing new questionnaires and accuracy (from 1 survey per person using 2 days to 1 survey per person using half hour).
3. **Internal Research Projects:**
   - Real-time Twitter sentiment in UK shopping centres: Connect to twitter API using python to access data and feed the processed data into a real-time topic modelling and sentiment dashboard using R Shiny.
   - Gravity model improvement: Investigate the feasibility of incorporating Agent-based Modelling using granular mobility data into existing gravity model.

EDUCATION

MSc Smart Cities and Urban Analytic, University College London *Sep 2017- Sep 2018*

London, United Kingdom

Applied advanced techniques to conduct spatial-temporal big data challenges.

Core modules: Quantitative methods; Geographical information systems; Spatial-Temporal Data Mining; Spatial capture and analysis; Urban simulation

BEng Civil Engineering, University of Shanghai for Science and Technology

Sep 2013 – Jun 2017

Shanghai, China

Core modules: Structural mechanics, Fundamental of structural steels, Structural Engineering

PERSONAL PROJECTS

COMMUN Ltd

Data Insight Scientist *London, United Kingdom*

Designed data model from back-end and created an insightful dashboard using Google Data Studio to facilitate decision-making related to projects funded by the community charities across the UK for the past five years. A classification model was built to assign topics for different funding projects based on features extracted from grants descriptions. ONS's national wide demographics data was used to find the correlation to identify where funding gaps my exist, what communities are in need of funding, pressing topics that could benefit from funding and how funding could be allocated more effectively, encouraging collaboration and avoiding the replication of efforts as a whole.

CERTIFICATES

- Machine Learning Engineering Udacity Nanodegree
- Alteryx Core & Advanced Certificate
- Certificate of Shanghai Interpretation Accreditation Test (Intermediate & Advanced Level)
- Dreamweaver web designer Certificate

TECHNICAL SKILLS

- R (4 years), Python (8 years), SQL (8 years), Git (3 years), PySpark (3 years)
- Treasure Data, AWS, GCP, Azure (2 years), Snowflake
- Machine learning algorithms and applications (8 years)
- BI Tools - QlikSense/QlikView/Tableau (4 years)
- Web design - Html5/CSS/JavaScript/Django/Leaflet/NodeJS (2 years)

LANGUAGES

- English (Proficient), Mandarin (Native), Shanghainese (Native)

FACTS ABOUT ME

- Dog lover with a penchant for hiking and travelling