

# IRFAN ALI KHAN, PhD, *(He/his/him)*

Senior Data Scientist | Adv Marketing Analytics & Revenue Optimization & Forecasting Expert | Economist |

 irfanaliguarulhos@gmail.com  +55 11 96932-8044 |  São Paulo, Brazil  
 [linkedin.com/in/irfan-ali-khan-93b52b159](https://www.linkedin.com/in/irfan-ali-khan-93b52b159) |  [Portfolio](#)

---

## PROFESSIONAL SUMMARY

**Senior Data Scientist and Analytical Researcher** with 10+ years delivering revenue growth through **marketing analytics, forecasting models, and predictive intelligence**. Ph.D. economist specializing in **Attribution Modeling, Customer Segmentation, Revenue Forecasting, and Distribution Optimization**. Expert in modern data stack (dbt, Snowflake, Power BI) and **ML/AI** (TensorFlow, PyTorch, XGBoost).

**Impact:** 18-23% marketing ROI improvements • \$4.5M+ revenue opportunities • 26% retention increases • 200+ analysts mentored

---

## CORE COMPETENCIES

**Analytics and Scientist:** *Marketing Mix Modeling (MMM) | Multi-Touch Attribution (MTA) | Customer Segmentation & CLV | A/B Testing | Revenue & Demand Forecasting | Distribution Modeling | Pricing Optimization | Churn & Retention Analytics | Campaign Performance | Time Series Demand Forecasting Modeling |*

**Data Science & ML:** *Python (Pandas, Scikit-learn, PySpark) | R | TensorFlow | PyTorch | XGBoost | LightGBM | Time-Series (ARIMA, Prophet) | NLP & LLMs (LangChain) | Deep Learning | Feature Engineering | Statistical Modeling | Econometrics*

**Data Engineering:** *dbt (Modeling, Testing) | Snowflake | Databricks | AWS (Redshift, Glue, Lambda) | Azure | GCP BigQuery | Apache Spark | ETL/ELT Pipelines | SQL (Advanced) | API Integration | CI/CD (GitHub)*

**Visualization & BI:** *Power BI (DAX, Semantic Models) | Tableau (LOD Expressions) | Looker | Plotly | Data Storytelling | KPI Dashboards*

**Tools & Platforms:** *Segment CDP | Braze | Mixpanel | SPSS | EViews | Excel (VBA, Power Query) | Jira | Agile/Scrum*

**Languages:** *English (Fluent) | Portuguese (Professional) | Spanish (Intermediate) | Hindi, Urdu, Pashto (Native)*

---

## PROFESSIONAL EXPERIENCE

### Marketing Advanced Modeling & Data Analytics Specialist

**Farmers Insurance (Capgemini) | Remote, USA | March 2025 – Present**

Lead advanced analytics for Direct Business Unit building **attribution models**, **customer segmentation**, **revenue forecasts**, and **predictive marketing models** driving multi-million dollar optimization.

**Key Achievements:**

- Designed **Multi-Touch Attribution (MTA)** and **Marketing Mix Models (MMM)** across 15+ channels, delivering 18% efficiency improvement and \$2.3M incremental revenue
- Built **ML customer segmentation** (K-means, RFM, hierarchical clustering) increasing campaign conversions by 22%
- Developed **churn prediction**, **CLV forecasting**, and **retention models** using XGBoost achieving 0.89 AUC scores
- Architected **A/B testing framework** enabling 50+ concurrent experiments monthly with statistical rigor
- Led **dbt transformation layer** on **Snowflake** creating 200+ reusable, tested data models with GitHub version control
- Engineered **Power BI semantic models** with optimized DAX serving 500+ users across business units
- Mapped **customer journeys** using **Segment CDP** and **Braze**, increasing retention by 14%
- Built automated **marketing dashboards** tracking ROAS, CAC, LTV:CAC, and attribution metrics
- Developed **revenue forecasting models** (Prophet, ARIMA) supporting quarterly planning and budget allocation

**Tech Stack:** *Snowflake* / *dbt* / *Power BI* / *Python* / *SQL* / *Segment* / *Braze* / *Mixpanel* / *GitHub* / *TensorFlow*

---

## Senior Data Scientist & Analyst

**Statistics Without Borders** | *Remote, USA* | **April 2024 – February 2025**

Led data science for global non-profit building **predictive systems**, **geographic analytics**, and **BI solutions** for educational programs across multiple US states.

**Key Achievements:**

- Managed **multi-state educational analytics projects** analyzing program effectiveness, identifying \$850K optimization opportunities
- Developed **ML forecasting models** (TensorFlow, Random Forest) for enrollment and resource allocation with 23% RMSE improvement
- Built **geospatial distribution models** examining regional disparities informing expansion across 15+ states
- Engineered **automated ETL pipelines** integrating 20+ sources processing 500K+ records daily with 99.8% accuracy

- Implemented **LangChain NLP solutions** for document classification and sentiment analysis achieving 91% accuracy
- Created **Tableau/Power BI dashboards** with predictive analytics enabling stakeholder self-service
- Led training for 15+ junior analysts accelerating project delivery by 40%

**Tech Stack:** *Python / TensorFlow / PyTorch / SQL / Tableau / Power BI / R / LangChain / ETL Pipelines*

---

## Data Scientist and Researcher (Forecasting & Adv Modeling)

**KPMG Australia** | *Remote, Australia* | **August 2022 – April 2024**

Delivered end-to-end analytics, forecasting, and optimization solutions for Fortune 500 clients across retail, finance, and technology. Specialized in building scalable predictive models, customer intelligence frameworks, and automated data pipelines that directly improved revenue, profitability, and operational efficiency

### **Key Achievements:**

- Analyzed 2M+ customers using **RFM analysis** and **clustering**, identifying \$4.5M revenue opportunities
- Built **time-series demand forecasting models** (ARIMA, Prophet) predicting quarterly revenue with 87% accuracy
- Developed **GLM/GAM pricing models** improving profit margins by 12% across key product lines.
- Created **geographic distribution models** optimizing logistics and reduced operational costs by 8%
- Designed **Tableau/Power BI dashboards** with custom KPIs serving 200+ business stakeholders
- Streamlined **ETL processes** reducing processing time by 45% maintaining 99.9% data integrity

**Tech Stack:** *SQL / Python / R / Tableau / Power BI / Databricks / Statistical Modeling*

---

## Data Analytics & Visualization Specialist

**Accenture** | *Remote, Brazil* | **January 2019 – July 2022**

Supported enterprise clients by delivering **end-to-end analytics**, ETL development, database management, and business intelligence solutions. Specialized in SQL-based data engineering, **customer segmentation**, **marketing analytics**, and **transforming complex datasets into clear business insights**.

### **Key Achievements:**

- Managed and optimized **SQL Server databases**, writing complex queries, stored procedures, and performance-tuned data models for analytics and reporting.
- Built and maintained ETL pipelines for data migration, transformation, and integration across **CRM, SAS**, marketing, and sales platforms.

- Performed **customer segmentation** using behavioral, demographic, and transactional data to support targeted marketing and retention strategies.
- Delivered business analysis, **translating raw data into insights** that improved decision-making across sales, marketing, finance, and operations.
- Built dimensional **data models that reduced reporting time** by 60% and improved data accessibility for business teams.
- Implemented **API-driven data pipelines** for payment and retail systems with automated validation and transformation.
- Built advanced Tableau dashboards using LOD expressions, earning an internal innovation award.
- Developed Power BI dashboards using DAX for KPI tracking, marketing analytics, and sales performance insights.

**Tech Stack:** *Python / SQL / R / Tableau / Power BI / Databricks / Statistical Analysis*

---

## **EDUCATION**

**Ph.D. in Economics** | *Universidade de São Paulo (USP)*, Brazil | **2020 – 2025**  
**Postgraduate Degree in Big Data** | *FACUMINAS*, Brazil | **Dec 2020 – May 2023**  
**M.Phil. in Economics** | *Preston University* | **2016 – 2018**  
**MBA in Accounting & Finance** | *Preston University*, Pakistan | **2011 – 2013**

---

## **CERTIFICATIONS**

- 🏆 Google Advanced Data Analytics Specialization (Big Query)
  - 🏆 IBM Certified Data Engineer – Big Data
  - 🏆 AWS Certified Data Scientist & Machine Learning
  - 🏆 Advanced Big Data Science & ML Certificate
- 

## **FEATURED PROJECTS**

**Multi-Touch Attribution Engine:** ML-based model analyzing 10M+ touchpoints across 15 channels distributing \$60M+ budget with 94% accuracy

**CLV Optimization System:** Deep learning model (TensorFlow) achieving 0.91 R<sup>2</sup> increasing LTV by 26%

**Real-Time Marketing Dashboard:** Snowflake + dbt + Power BI solution processing 2M+ daily events with <5 min latency

**Demand Forecasting Engine:** Ensemble system (ARIMA, Prophet, XGBoost) achieving 87% accuracy across 500+ SKUs

**Revenue Optimization Model:** Dynamic pricing algorithm increasing revenue by 15%

**A/B Testing Platform:** Python framework automating experimental design reducing analysis time by 75%

---