

Malavika Mampally

+1 9192363489 | malavikamampally@gmail.com | linkedin.com/in/malavikamampally | github.com/malavika-vm

SUMMARY

Data Scientist with an MS in Statistics and hands-on experience in healthcare analytics, patient segmentation, statistical modeling, and commercial performance measurement. Skilled at translating complex analytical outputs into actionable insights for cross-functional stakeholders. Proficient in Python, R, SQL, and Tableau with experience applying ML and advanced analytics to large-scale datasets in healthcare and insurance contexts.

SKILLS

Data Analysis and Visualization Tools: Python (pandas, scikit-learn, PyTorch, seaborn), R (tidyverse, Shiny, ggplot2), SQL (query optimization, data extraction and transformation), MATLAB, Tableau, Power BI, Excel, Git, Docker
Analytics and Machine Learning: Regression (lasso, ridge, PLS, PCA), Classification, Clustering, Patient/Market Segmentation, Forecasting, A/B Testing, Feature Engineering, Model Validation, Exploratory Data Analysis, Statistical Modeling, Performance Analytics, Dashboard Development and Reporting

EXPERIENCE

Research Associate - Data Science Society-Centered AI Lab, UNC, <i>Chapel Hill, NC</i>	01/2025 – 12/2025
<ul style="list-style-type: none">Served as the lead statistician for a large-scale benchmarking study analyzing healthcare survey response patterns to evaluate AI deployment readiness in sensitive health topicsApplied machine learning techniques to simulate and model LLM responses to sensitive health topics, performing exploratory data analysis, feature engineering, and model comparison to assess bias patternsCo-authored research paper on <i>Measuring Multi-level Human-LLM Alignment</i>; preprint available on ArXiv (https://arxiv.org/abs/2512.13142), manuscript under peer review	
Data Science Consultant NC TraCS Institute, UNC Health	01/2025 – 05/2025
<ul style="list-style-type: none">Built the first end-to-end analytics framework for a patient recruitment system, transforming an unmonitored process into a repeatable analytics workflow with interactive dashboards and data visualizationsReduced delays by 40% by conducting descriptive and longitudinal data analysis of REDCap data to track ticket completion trends and identify systemic bottlenecks that informed commercial decision-makingDemonstrated 285% improvement in operational strategy adoption by applying Fisher's Exact Test, delivering actionable insights to clinical stakeholders that informed resource reallocation decisions	
Statistical Consultant Dept. of Exercise & Sports Science, UNC, <i>Chapel Hill, NC</i>	01/2025 – 05/2025
<ul style="list-style-type: none">Evaluated a fitness program by conducting performance analytics across 180+ fitness variables and 1,900+ students to inform data-driven program design decisionsDesigned A/B tests across participant segments (age, gender, program type), defined KPIs, and built automated dashboards and reports to surface actionable insights and support ad-hoc analyses	
Data Analyst Intern Future Generali India Insurance Co., <i>Mumbai, India</i>	06/2022 – 08/2022
<ul style="list-style-type: none">Reduced reserving data processing time by 30% by streamlining quarterly workflows across 10+ insurance lines, managing large-scale commercial datasets (100K+ records) with validation checks that ensured data accuracy and compliance for pricing and reserving modelsIdentified key risk factors driving motor insurance premium variation by performing regression analysis on policy-holder data using R, producing findings that informed enhanced pricing strategies for the underwriting team	
Graduate Teaching Assistant School of Data Science and Society, UNC, <i>Chapel Hill, NC</i>	08/2023 – 05/2025
<ul style="list-style-type: none">Tutored students in 100-level Data Science and Statistics courses and mentored 10+ project groups on data analysis methodology and statistical modeling	

PROJECTS

Life Expectancy Prediction Model (R): Built end-to-end predictive model for country-level health outcomes from WHO data using LASSO, Ridge and PCA regression with full model validation pipeline to support forecasting inputs

NBA Game Spread Prediction (Python) — Engineered novel features and built ensemble ML models achieving 96% accuracy; ranked #1 among 14 teams for lowest prediction error

EDUCATION

MS in Statistics, Analytics & Data Science University of North Carolina at Chapel Hill	05/2025
High Pass (>90%). Coursework: Machine Learning, Applied Statistics, Stochastic Modeling, Sports Analytics, Design Thinking	
BS in Statistics University of Mumbai	05/2023
9.69 GPA PG Diploma in Applied Statistics (Valedictorian) Actuarial Exams: P (Probability) & FM (Financial Mathematics)	