



Pratik Majumdar

Nationality: Indian **Date of birth:** 25/08/2004 **Place of birth:** Kolkata

Gender: Male **Phone:** (+91) 8617262687 **Email:** pratik.m2508@gmail.com

LinkedIn: [linkedin.com/in/pratikmajumdar/](https://www.linkedin.com/in/pratikmajumdar/)

Website: <https://pratikmj004.github.io/>

Home: Kolkata (India)

ABOUT MYSELF

A budding statistician, who takes deep interest in the various application of statistical methods, particularly in the domain of Biostatistics and Epidemiology. As a fast paced learner with inquisitive mindset, I seek to know how and which statistical concepts can be handy to solve several issues relating to public health.

SKILLS

Julia / R, RStudio / Python / C / Google Suite (Doc, Slides, Form, Sheet, Drive) / Basic SQL (MySQL) and Tableau Software / Power BI / biostatistics / epidemiology / Machine Learning / Time Series Analysis / Data Analysis (Inferential Statistics, T-Test, Correlation Coefficient) / statistics / Probability Theory / Stochastic Processes / Microeconomics / Macroeconomics / Public Finance

EDUCATION AND TRAINING

MS in Statistics

Presidency University [15/09/2025 – Current]

City: Kolkata | **Country:** India | **Website:** <https://www.presiuniv.ac.in> | **Field(s) of study:** Natural sciences, mathematics and statistics: • Statistics | **Level in EQF:** EQF level 7

BS in Statistics with Honors

Presidency University [12/09/2022 – 01/07/2025]

City: Kolkata | **Country:** India | **Website:** <https://www.presiuniv.ac.in> | **Field(s) of study:** Natural sciences, mathematics and statistics: • Statistics ; Social sciences, journalism and information: • Economics | **Final grade:** A (7.84/10) | **Level in EQF:** EQF level 6 | **Type of credits:** Semester System | **Number of credits:** 120

LANGUAGE SKILLS

Mother tongue(s): Bengali

Other language(s):

English

LISTENING C2 **READING** C1 **WRITING** B2

SPOKEN PRODUCTION B2 **SPOKEN INTERACTION** B2

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user

PROJECTS

[01/06/2025 – 30/06/2025]

Medical Insurance Charges Prediction

- **Project Focus:** Investigated determinants of medical insurance charges for 1,338 patients using demographic and health data. Developed and evaluated multiple regression models using an Ordinary Least Square approach to predict annual charges
- **Actions:** Conducted Exploratory Data Analysis (EDA) and applied Statistical Hypothesis Tests (Levene's & Chi-Square) to confirm variable significance. Employed OLS regression in R, including a 70/30 data split for model training and validation. Evaluated model performance using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).
- **Key Findings:** Smoking status, BMI & Age, were most significant predictors. *Was able to explain about 76% of the total variability in the insurance charges by a parsimonious model.* Quantified that smokers incur \$24,069 more in insurance charges, as compared to the non-smokers.
- **Recommendations:** Proposed policy interventions to focus on smoking restrictions and healthy living to manage premium costs. Suggested future work to incorporate prior medical history as a regressor and apply advanced ML models (e.g. Random Forest, Gradient Boosting, etc.), for improved precision.

Link: <https://tinyurl.com/svpf27yd>

[01/07/2025 – 31/07/2025]

Survival Analysis for Lung Cancer

- **Project Focus:** Modeled survival outcomes for 228 advanced lung cancer patients, analyzing the influence of clinical and nutritional factors (e.g. ECOG, Karnofsky scores, calorie intake, etc.). Developed a Cox-Proportional Hazards Model to predict patients survival time.
- **Actions:** Applied core Survival Analysis techniques, including Kaplan-Meier Plot and Log-Rank Test to assess survival probability differences between patient groups. Developed the predictive model using the Cox-Proportional Hazards model in R, including a 70/30 data split for model training and validation. Assessed the model's discriminatory power using the Concordance Index (C-index).
- **Key Findings:** Confirmed the Sex, and patient's ECOG scores are confirmatory predictors of a patient's demise. The Log-Rank test showed that Females have a statistically significant better survival probability as compared to Males ($p\text{-value} = 0.01$). *The Cox model demonstrated good discriminatory power with a C-index of 0.726.* Found that Karnofsky scores (physician/patient rated) were statistically insignificant in predicting survival time.
- **Recommendations:** Proposed closer clinical monitoring of less significant factors like Calorie Intake to improve patient care, noting a higher percentage of male died with below-average intake. Suggested future work to utilize a larger sample size and explore advanced Survival analysis methods to obtain more discriminative Cox models.

Link: <https://tinyurl.com/8btfp7ct>

VOLUNTEERING

[12/09/2022 – Current] Kolkata

National Service Scheme

- Went across slums to provide with preventive measures against Dengue.
- Participated in tree plantation drive in the campus.
- Visited numerous schools to promote child safety and concerns.
- Secured National level prize as a team, for numerous nature conservative drives.