

**Ushnik Mukherjee, PhD**  
**Kitchener, ON, N2M 1B4 Canada**  
mukherjee.ushnik@gmail.com  
+1-5197814159

**SUMMARY**

Highly motivated applied science data scientist with extensive experience in developing and implementing data pipelines, optimization models, and spatial analytics in the energy, transportation and agricultural sectors. Proficient in Python and SQL, with a proven ability to process complex datasets and deliver actionable insights. As a boyhood football fan, I am particularly eager to transition into a Data Engineer role and leverage my strong analytical and technical skills to contribute to the development of cutting-edge data infrastructure for football analytics.

**EXPERIENCE**

---

*UNIVERSITY OF WATERLOO / WATERLOO, ON*

Jan 2022 -  
Present

**Research Associate (Data Science & Optimization, Geospatial Analytics)**

- Developed a Python ETL pipeline to extract and clean data from Statistics Canada's online database, removing text annotations and preprocessing animal count and employment data for accurate analysis
- Utilized a data disaggregation approach, combining census data with satellite imagery and geospatial datasets, to generate high-resolution spatial Pan Canada data. This involved downscaling animal counts to barn locations and allocating food waste to urban areas, showcasing strong spatial data processing and integration skills
- Prepared data for model inputs through rigorous cleaning and preprocessing, formulating linear programming mathematical optimization models to determine optimal waste management policies through biodigesters
- Delivered data-driven insights to conservation authorities, energy firms, and policymakers, supporting resource allocation and environmental planning

Jan 2021 -  
Jan 2022

**Parental Leave (1 year)**

Oct 2018 -  
Jan 2021

**Postdoctoral Fellow (Transportation Data Analytics & Policy Evaluation)**

- Engineered a Python data pipeline to process traffic simulation outputs, assigning zonal markers to trips and integrating GIS road network data. This pipeline generated inputs for the US EPA's MOVES emissions simulator
- Analyzed large datasets from the MOVES emissions simulator using SQL, quantifying the environmental impact of high-emission freight routes and demonstrating strong relational database skills
- Created spatial visualizations using Python libraries to communicate aggregated emission data, enabling stakeholders to identify high-impact freight corridors

Sept 2013 -  
Aug 2018

*GRADUATE RESEARCH ASSISTANT / WATERLOO, ON*

**PhD Candidate (Data Analytics & Optimization, Energy Market Analysis)**

- Extracted and prepared electricity market data (pricing, generator dispatch, and demand) from the Independent Electricity System Operator's database
- Developed optimization models to generate optimal energy storage system dispatch signals, demonstrating the application of algorithms to solve complex problems
- Analyzed model results using time series analysis to evaluate and refine energy storage dispatch algorithms, providing insights for improved dispatch strategies
- Explored stochastic modeling of electricity price and demand by gathering historical data, organizing it by hour of day and season over a 10-year period, fitting probability distribution functions, generating possible realizations, and modeling a stochastic programming optimization model for energy storage dispatch
- Designed and managed data storage, utilizing CSV formats, to efficiently store and retrieve large volumes of electricity market data

## EDUCATION

---

Aug 2018

### **Doctor of Philosophy in Chemical Engineering**

*UNIVERSITY OF WATERLOO / WATERLOO, ON*

- Thesis: Optimal Planning and Operation of Power to Gas Energy Hubs in Transitioning Electricity Systems
- Coursework: Statistics in Engineering, Applied Engineering Mathematics, Design Optimization Under Uncertainty

## SKILLS

---

- Programming & Software: Python (NumPy, pandas, scikit-learn, Plotly), SQL (querying & filtering)
- Data Science & Machine Learning: Statistical modeling (regression, time series), machine learning (logistic regression, PCA), optimization (linear programming, stochastic), and geospatial analysis
- Data Visualization & Web Development: Interactive dashboards (Plotly)
- Database Management: Cleaning, processing, and analyzing large, structured datasets
- Communication & Collaboration: Presenting data insights to technical and non-technical stakeholders

## HOBBIES

---

In my free time, I work on data science projects related to football analytics and look at different mathematical approaches to understand the game. I maintain a github webpage of the projects I have worked on so far here: <https://u3mukher.github.io/x-stats/blog/>

### **Developing the Expected Disruption (xD) Model: Valuing Defensive Actions Solely on Event Data**

*StatsBomb Data: Top 5 European Leagues, 2015/16*

- Developed an xD model to quantify defensive impact, expanding on xT methodology
- Created interactive Plotly heatmaps for visualizing disruption patterns
- Ranked top performers by pitch thirds using possession-adjusted xD

### **Principal Component Analysis (PCA) for Football Transfers**

*StatsBomb Data: Top 5 European Leagues, 2015/16*

- Applied PCA on 13 performance metrics to profile playing styles of players and teams
- Ranked players and teams across top 5 leagues based on percentiles for gauging strength
- Clustered players and teams based on attack, possession, defending, and set piece
- Evaluated player-team fit, identifying if transfers made sense

### **Analyzing Positional Contributions with VAEP**

*StatsBomb Data: Bayer Leverkusen, 2023/24*

- Trained an XGBoost-based VAEP model on event data
- Measured on-ball contributions across defensive, midfield, and attacking units by match categories: Comfortable Wins, Close Wins, and Late Equalizer Draws for Leverkusen's unbeaten season
- Created strip plots and cumulative VAEP charts for match impact analysis

## Professional References

---

- Nandita Basu, Professor, University of Waterloo
- Juan Moreno-Cruz, Professor, University of Waterloo