Statement of purpose

My journey into Data Science began through exposure to real-world applications during conversations with my friend, a Data Engineer at Amazon, on whose account I understood what enables predictive algorithms to predict market trends that influence global decisions and how data analysis transform business decisions with insights. It fascinated me how predictive algorithms help revolutionize customer recommendations and in doing so the world around us. This fascination evolved into a deeper appreciation for the underlying mathematical foundation through exploration of Youtube Channels like 3Blue1Brown, StatQuest, and particularly Steve Brunton, whose work inspired me to pursue an higher education in Data Science. Further insights from my uncle, a VP at Moody's illuminated how statistical methodologies in finance seamlessly integrate with ML algorithms, ultimately pushing me to pursue advanced studies in this rapidly evolving field.

Building upon this motivation, I systematically developed my technical foundation through structured learning. I began with Machine Learning Onramp and Deep Learning Onramp by MathWorks, progressing to R Programming for Data Science on Udemy to gain hands on experience with statistical computing. To understand the mathematical foundation of LLMs and other ML models, I started with Gilbert Strang's "Linear Algebra and Its Applications", and to bridge the gap between theoretical concepts with practical ML applications I completed the "Mathematics for Machine Learning: Linear Algebra" on Coursera by Imperial College London. Currently I am learning my way around the ropes on training and fine-tuning LLMs and ML models using CUDA, PyTorch and running them on Ollama. I'm also trying to create AI agents and orchestrate a RAG pipeline using tools like n8n and langflow. Additionally I'm trying to add to my learning of foundational concepts by learning Multivariate Calculus and Statistics.

To bridge theory and practice, I developed a global renewable energy analysis using real-world data from Our World in Data. This project involved statistical modeling in R, achieving R^2 = 0.89 predictive accuracy for renewable energy forecasting. The experience demonstrated my ability to handle complete data science workflows, from data collection through statistical analysis to insight generation. This experience was transformative, requiring me to implement statistical models, and present complex data relationships clearly. It reinforced my passion for using data science to address global challenges, particularly in sustainability and climate policy with real-world applications.

The Master of Data Science at the University of Melbourne represents the ideal environment for advancing my technical and mathematical expertise. I am drawn to the opportunity to learn from professors like Howard Bondell, whose extensive research portfolio and high-impact publications demonstrate the calibre of statistical expertise I aspire to have. His work on simultaneous regression techniques represents the kind of statistical modeling I hope to explore beyond my current linear regression foundation. The program's core subjects in Statistical Machine Learning and Computational Statistics will help me employ more sophisticated methods and regularization techniques. Melbourne's Melbourne Data Analysis Platform (MDAP) and collaboration with the Vector Institute further demonstrate the university's commitment to remaining at the forefront of Data Science education.

Upon completing this program, I aim to pursue a career as a Data Scientist specializing in environmental analytics and sustainability consulting. My immediate goal after graduating is to join leading consulting firms where I can apply advanced machine learning techniques to global energy transition strategies. The interdisplinary nature of Data Science appeals to me as it can be coupled with any sector to address global challenges. I hope to develop predictive models for renewable energy integration and sustainable development.