

Maciej P. Zieliński

Recent Data Science Graduate

Phone: (+48) 503143080
(+44) 7510081053

Email: izielak@me.com

LinkedIn: maciejzielinski65537

Github: bajun65537

Portfolio: maciejdrawsdata.com

Short bio

I am a recent graduate looking for the real-world experience in fields of data analytics and machine learning.

I have read for MA in Economics at the University of Glasgow, my alma mater. For my postgraduate studies I chose the University of Edinburgh from which I have recently graduated with MSc in Operational Research and Data Science.

I also enjoy solving puzzles, coding and maths. I am passionate about data and its numerous applications. In addition to being a data-focused individual with a conventional mathematical intuition, I am an effective communicator who can translate complex solutions to less technical audience.

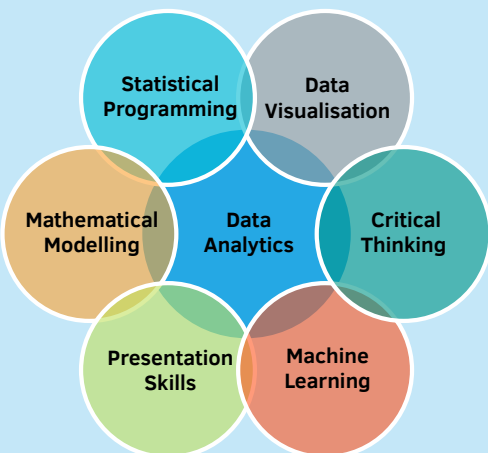
I constantly develop my skills and try to share my knowledge with others. For the purpose of data visualisation, data analysis or machine learning I mostly work with Python and R.

My academic interests include:

- Computational Geometry
- Mathematical Modelling
- Optimisation
- Environmental Economics

Skills

Overview



Education

2020 - 2021 **MSc, Operational Research and Data Science** University of Edinburgh

- **Operational Research:** Studied dynamic programming for sequential decision problems and integer programming with logical constraints. Solution methods included the Branch-and-Bound and Gomory Cuts. Solved a number of applied problems from various fields like: Energy, Aviation, Healthcare, Sport or the Environment.
- **Optimisation:** Covered the mathematical theory behind the Simplex method, studied the role of duality and sensitivity analysis for solving linear optimisation problems.
- **Simulation:** Generated variates from standard distributions including normal, gamma, beta, Poisson, binomial or Weibull. Used the Arena and Simul8 packages to build and experiment on models in various manufacturing and management systems.
- **Machine Learning:** Studied the theory behind a number of machine learning methods and applied them in practice using Python.
Familiar models: Lasso/Ridge regressions, logistic regression, naive Bayes, maximal margin classifiers, support vector machines, decision trees, random forests, k-means or hierarchical clustering.
- **Statistics and Probability:** Covered topics such as maximum likelihood estimation, likelihood ratio tests, Bayes theorem and posterior distribution, Fisher's method of scoring, generalised linear models. Used R to fit generalised linear models.
- **Missing Data:** Analysed various datasets with missing data. Focused on single and multiple imputation methods as well as likelihood based EM algorithm and their implementation using R.
- **Energy Markets:** Combined OR theory with modelling techniques to study a range of current energy industry challenges including: optimal power flow, unit commitment, power system planning.
- **Credit Scoring:** Studied scorecard development processes using both statistical and non-statistical methods including regressions, clustering, neural networks or genetic algorithms.

2016 - 2020 **MA, Economics** University of Glasgow

- **Econometrics:** Analysed different economic, financial, political and social datasets. Covered various models including multivariate OLS, instrumental variables, panel data, time series and some other non-linear models.
- **Game Theory:** Applied game theoretic methodology to model the behaviour of economic agents interacting in markets or during elections.

Jun 2017 **Summer School Programme** University of Nottingham Ningbo China

Covered topics such as China's economy, how to start and manage a social enterprise in China, the corporate social responsibility in China and the growing innovation in China. Moreover, I studied Mandarin with the focus on a professional engagement in China.

Thesis

MSc, Operational Research and Data Science Grade: **A2**

"Facility Location and Road Design."

This dissertation builds upon existing literature on facility location problems that include optimal positioning of the rapid transit line inside the rectangular Manhattan-like city centre. I presented a bi-objective model for determining the location of a motorway. The objectives are to minimise the total closest distance and the expected travel time. The closest distance describes the accessibility of the customers, while the expected travel time measures the impact the motorway has on a time distance.

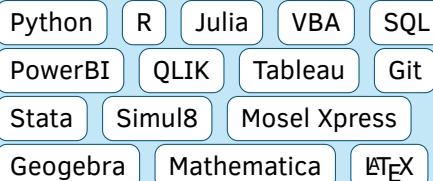
MA, Economics Grade: **A1**

"Carbon tax hits the road. Evidence from British Columbia transport sector."

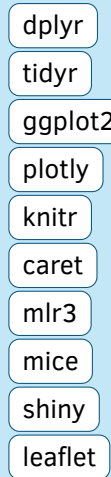
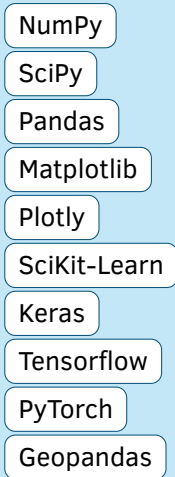
This dissertation builds upon existing literature on carbon taxation effectiveness in mitigating climate change. In this quasi-experiment study I analysed British Columbia carbon tax policy and its impact on transport sector emissions. I have built two models using Difference-in-Difference and Synthetic Control Method in order to estimate the possible effect.

Software

I am familiar with a number of tools most commonly used in statistical programming, data visualisation, machine learning, creating dashboards, process automation and optimisation.



In particular, for the purpose of data analysis, visualising data and machine learning I have mostly used Python and R. Below I have shortlisted some of the libraries and packages I am familiar with.



Finally, **Adobe Illustrator** enables me to further expand on the visualisations made in R or Python and make them even more aesthetically pleasing.

Professional Development

Outside my degree I constantly tried to broaden my knowledge. I have already obtained the following certificates via HarvardX:

- R Basics
- Data Visualisation
- Probability
- Inference and Modelling
- Productivity Tools
- Data Wrangling
- Linear Regression
- Machine Learning
- High-Dimensional Data Analysis

Similarly, I have audited via IBM:

- Deep Learning with Keras
- Deep Learning with Tensorflow
- Deep Learning with PyTorch

Experience

Jul 2019 -
Sep 2019

Process Automation Intern

Prinzhorn Group - Dunapack Packaging

During the internship I was offered a variety of learning opportunities among different departments including HR, production planning, sales analysis, transportation, controlling and IT. Being exposed to multiple roles and perspectives gave me an inside glimpse of how the packaging industry operates. After being briefed of what daily, weekly or monthly responsibilities are in each department, I tried to understand which tasks are repetitive-type and how one could try to automate them. Tools that I have been working with during the internship were: VBA, Windows Task Scheduler, Power Bi, QLIK or the SAP ERP.

Jul 2018 -
Sep 2018

Data Analysis Intern

Amcor

I gained experience in using Excel for various purposes (data handling, data presentation, building regression models and presenting findings of my analysis). In order to facilitate some of the processes I used pivot tables, array formulas, macros and VBA. What is more, I understood the limitations of Excel, especially lagging when exceeding human-digestible levels of data and the cumbersome CPU usage when applying formulas that recalculate a large number of cells.

During my internship within the Finance Department I was able to:

- Perform various regressions (OLS, GLM, robust) on a real world data to make predictions about the future.
- Deal with imperfect datasets.
- Get involved in modelling of cost savings and process mapping.
- Contribute towards creating financial reports and analysis.

Projects

• What makes a TV series good?

[More information available here.](#)

Aim: The purpose of the project was to advise NBC Universal how to produce the best and highest rated reunion episode possible.

Action: In order to produce an episode that viewers would love, I needed to find out what made some episodes more popular than others. The task was to use data to build a predictive model that would capture the underlying relationships between features and the IMDb ratings.

- Performed an exploratory analysis of data from `SchrutePy` package to see which features may or may not have value in terms of predicting the IMDb rating of an episode.
- Using external sources I was able to find a dataset containing the entire show's transcript line by line. The idea was to perform a text analysis to see whether there are particular jokes which happen to be better received by viewers and therefore, result in more popular episodes.
- After that, I have prepared the chosen variables for the modelling stage via the process of the feature engineering.

Results: Throughout modelling stage I have identified a number of key features which might be crucial in making a great reunion episodes. Having explored a number of models, e.g. linear, Ridge and Lasso regressions, I have decided to use Lasso model as it had the lowest root mean squared error.

Keywords: data analysis, data visualisation, Ridge/Lasso regression

• Scottish Road Network Exploration

[More information available here.](#)

Using the Department of Transport dataset which contains counting points information allows to analyse some interesting patterns of British traffic and the UK's road network.

Aim: This project focuses on the analysis of geographical features of the Scottish Road Network and its usage.

Results:


- By looking at the different sections of the Scottish Trunk Network I found out which section has the most traffic and which mode of transportation is the most prominent.
- Found the busiest roads in Scotland and their location.
- Analysed changes in traffic during the day for M8 road. Captured trends in people commuting between Glasgow and Edinburgh.
- Explored the question whether the longest roads coincide with the busiest roads.
- Looked at which local authority manages the largest number of roads.

Keywords: data analysis, data visualisation, geographical data, Geopandas


Languages

A1  C2

Polish (native)



English (proficient)




Mandarin (pre-intermediate)



German (elementary)



Spanish (basic)



Soft skills

Communication: I adequately convey complex ideas and data-based insights to a non-technical audience. Business case competitions enabled me to gain experience in public speaking.

Problem-solving: I always try to weigh pros and cons of each solution in order to find the path that will allow project to run smoothly.

Organisation: I effectively manage my tasks and divide the work accordingly to meet the deadline.

Leadership: I recognise organisations' objectives and their market position to provide better solutions to the current business problems. In addition, I keep the right distance from a task to see the bigger picture and take initiative whenever needed.

Teamwork: I try to embrace teamwork and share ideas with others to improve performance. I have collaborated with others when solving business cases or during numerous university projects.

Creativity: I constantly try to find new ways of looking at things and develop innovative project solutions. Thinking outside of the box helps me to solve more complicated problems as well as deal with unexpected errors.

Interests

Forums

I am an active member of the Kaggle and the Stack Exchange online forums. In my free time I enjoy reading kernels of fellow Kagglers and ask or answer questions posted on Stack Exchange.

Hobbies

I enjoy solving Diophantine equations, an occasional game of chess and lately, I have been interested in refurbishing old consoles and retro gaming.

• Will a hotel reservation be cancelled?

[More information available here.](#)

People tend to book their hotel rooms in advance - yet, they might cancel just a few days before their reservation causing hotels to lose their revenue. This is the reason why predicting which customers are more likely to cancel their bookings is so vital.

Aim: The purpose of the projects is to help the hotel management make more informed decisions with regards to safeguarding against lost revenue.

Action: The idea was to use provided data to build a predictive model that would let us capture the relationship between the booking information and whether the booking was cancelled.

- Performed the exploratory data analysis which included cleaning the data, dealing with any missing values and disregarding any observations or variables which were erroneous.
- Used the logistic regression as a baseline model.
- Further, explored decision trees and random forests.
- Assessed models' performance using the receiver operating characteristic and false positive rate.

Results: Despite logistic regression being slightly less accurate overall than some of the other models explored, it did give the lowest false positive rate. For this problem, false positives can lead to potential double booking and thus can be especially harmful for the hotel's reputation.

Keywords: data visualisation, logistic regression, decision trees, random forest

• Optimal Power Flow with Electricity and Heat Stores

[More information available here.](#)

Aim: In this project I modelled the electricity network with electricity and heat storage over multiple periods.

Action: For the purpose of this project I have used the DC approximation to model the power flow.

- Assumed that line resistances to be negligible compared to the line reactance in order to simplify the parameters and normalised voltage magnitudes.
- Declared sets of indices, introduced parameters and decision variables used within the model.
- Considered the effect the electricity production has on the environment and hence, included the carbon tax in the model.
- Made sure the physical constraints including Kirchhoff's circuit laws hold.
- Modelled the operating cost to be given by the cost of power generation from each of generators and the carbon tax multiplied by tonnes of CO₂ emissions released during the process of energy generation.
- Included the possibility of storing heat and electricity in the model.
- Implemented the model in Mosel Xpress and minimised the operating cost.

Results: Experimentally checked that if I were to duplicate line with the same reactance as the existing line there would be no advantage in having a higher capacity in the new line than in the existing line.

Keywords: operational research, mathematical modelling, optimisation

Achievements

• University of Edinburgh School of Mathematics Scholarship 2021

• 2nd place in Economic Futures Essay Competition 2020

[More information here.](#)

In a paper called "Covid-19 pandemic, its impact on air quality and possible long-run effect on the environment." I have empirically confirmed a short-run improvement in air quality using data on the amount of nitrogen dioxide in the air across various UK cities. Having visualised the global emissions from 1960 to present, a certain trend became transparent. After each huge event affecting worldwide economy, there was a small decrease in emissions. However, in the following years we could have observed emissions raising at even higher rate. Finally, I warned the reader that we should avoid boosting the economy in a way that will replace health crisis with yet another one - the amplified air pollution. All visualisations were made using R.

• 1st place in ASES & GUCS Business Case Competition 2020

From the pool of over 30 teams, we got selected and had a chance to compete in the finale. Our team, Penguin Solutions, have presented business solutions in various real world scenarios to a panel of judges representing consulting firms (e.g. Capco), academics and a wide audience. After the presentation we had a chance to answer various questions asked by the panel and clarify any ambiguities that might have arisen during the presentation.