PASCAL SCHMIDT

I am currently working as a data analyst, building dashboards, testing hypothesis, and doing all different kinds of data analysis. The main tools I use are R, Shiny, Python, and SQL. When I am not working my day job, I am blogging at thatdatatho.com and building web applications on my personal website about data that interests me.

EDUCATION

2014 | 2019

B.S., Statistics, (Minor Economics)

Simon Fraser University, in-major GPA: 3.55 — Burnaby, BC

WORK EXPERIENCE

November 2020 | Current

Data Analyst

LBC Studios — Vancouver, BC

- Investigated returns on bundle purchases by testing multiple hypotheses and recommended different price points that led to increased revenue for the analyzed bundles.
- Provided dashboards and reports to the marketing team to better understand marketing campaigns and revenue flow.

June 2020 | November 2020

Data Analyst

SAM — Edmonton, AB

- Deployed and dockerized a web application on AWS with R Shiny that analyzed real-time Twitter data, which led to increased productivity of employees who used the application for research purposes.
- Queried, cleaned, and visualized data with the help of `pandas`, `NumPy`, and Python plotting libraries to update teams about machine learning model performances.
- Developed automated interactive dashboards with `plotly` and `flexdashboard` to inform teams about the internal data usage and to improve the client experience by sending out quality information.

May 2019 | August 2019

Data Analyst, co-op

Statistics Canada — Ottawa, ON

- Developed R scripts for cleaning multiple data sets and used SQL joins to build a usable data set for analysis.
- Built an R markdown report with tables and graphs for a better understanding of why people's addresses on administrative data match that on census data.
- Created data visualization plots in Power BI to supplement the R markdown report.
- Presented data and conclusions to my team, which used the results to gain actionable insights into the Census Program Transformation Project and improved the population coverage for admin data.

May 2018 | December 2018

Data Science Trainee

BC Cancer Agency — Vancouver, BC

- Designed reports and conducted an explanatory data analysis, which led to a successful publication, by visualizing the key characteristics of patients with endometrial cancer.
- Produced interactive dashboards to communicate data to physicians who used my work for presentations.
- Independently researched and applied previously unknown statistical libraries for improved data representation.
- Composed a bi-annual report by cleaning, reshaping, and joining multiple unstructured data sources, working extensively with R tidyverse packages such as `dplyr`, `ggplot`, `purrr`, `tidyr`, and `stringr`.
- Implemented a meta-analysis of individual patient data and a cox-regression model to determine if endometrial cancer patients need additional treatment plans and to improve their quality of life.

CONTACT

pascal.sfu@gmail.com
Pascal-Schmidt
Pascal-Schmidt
thatdatatho

SKILLS

- R/RStudio
- Shiny
- Markdown
- Python
- C++
- SQL
- Power BI
- Docker
- AWS
- Version Control
- Statistical Analysis

COMMONLY USED LIBRARIES

`tidyverse`
`tidymodels`
`shiny` + `shinyjs`
`flexdashboard`
`Rcpp`
`pandas`
`NumPy`
`scikit-learn`

RESEARCH

Molecular subtype not immune response drives outcomes in endometrial carcinoma

PROJECTS

January 2018 | present

A Data-Driven Approach to Evaluating the Vancouver Housing Market

Tools Used: Shiny, Bootstrap, CSS, Docker, Docker Compose, MongoDB, AWS, shinyjs, leaflet, tidyverse, tidymodels, selenium, random forest

- Scraped multiple real estate websites for property prices and features and created a cleaned, standardized, and tidy data set for analysis.
- Fitted a random forest model with the `tidymodels` package that predicted home and rental prices and calculated price to rent ratios to inform Vancouverites about the best properties.
- Improved the random forest model (decreased the MAE by 20%) by augmenting the data set with new feature engineered variables by leveraging the Google Maps API.
- Created an interactive web application that serves as an exploratory tool and lets users use the machine learning model to make predictions about any desired property in Vancouver.
- Deployed the application with the help of Docker on AWS and made it available to the public.

Twitter Exploration Tool

Tools Used: Shiny, CSS, Bootstrap, DataTable, leaflet, plotly, shinyjs, rtweet

- Developed a web application that lets users explore the usage of words on Twitter for specified cities around the world.
- Implemented a network graph that visualizes bigrams from Tweets and created word clouds for handles, hashtags, and mentions to explore how people talk about topics on Twitter.

Matrix Completion of Weather Data

Tools Used: R, Rcpp, Python, pandas, numpy

- Won the in-class Kaggle competition by exploring NA values in the data set and predicting the missing values with a combination of linear interpolation and linear regression.
- Implemented a function in C++ that identified large gaps of missing values between observations and helped me choose between interpolation and regression.
- Decreased the run time of my notebook by 80% by re-writing some code blocks in C++.

Personal Blog

Tools Used: Passion to learn and teach data science and programming

- Published blog posts about programming in R and Python, using and exploring specific libraries, developing web applications with shiny, building predictive models, and statistics and data science concepts.
- Learned and internalized data science and programming principles better by teaching readers about concepts I am learning about.

PROFESSIONAL DEVELOPMENT

Tutoring

- Tutored students in programming and statistics classes to help students achieve their learning and grade goals and to build a strong foundation of these concepts for myself.

Kaggle Meet Ups

- Visited the Kaggle meet-up group and learned about competitions, algorithm implementation, and data science concepts.

Online Classes

- Completed online classes on platforms such as Business Science, Coursera, and Data360 to learn new technologies and expand my horizon as a data scientist.

#TidyTuesday

- Participated in weekly challenges that consisted of a data set, posted by R Studio for the R community, to analyze and then share the findings and analysis with the community on Twitter.