

Agnes Yunhong He

Data Science and Data Analytics

Yunhong@umich.edu cell: 647-608-3939 618-99 Cameron Street, Toronto, ON, Canada
<https://www.linkedin.com/in/agneshe/>

PERSONAL SUMMARY

I am passionate about solving business problems using Data Science & Machine Learning. I systematically & creatively use my skillset to add tangible value to the team, the business, and the end-user. I am constantly learning, and always looking to improve.

SKILLS & TOOLS

- Programming: SQL(10 yrs), Python (2 yrs, Base, Pandas, Numpy, Matplotlib, Scikit-learn), VBA (5 yrs)
- Tools: Tableau(2 yrs), Power BI(2 yrs), Excel(10 yrs, DAX, VBA macros, Power Query, data models, PivotTables, Formulas such as sumproduct, Sumif, countif, vlookup...etc.), Github (1 yr), AWS (1 yr, S3, Lambda, IAM, EC2, SageMaker, RDS, DynamoDB), MS Access(4 yrs)
- Statistics (2 yrs): Hypothesis Testing, AB Testing, Central Limit Theorem
- Machine Learning (2 yrs): Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, K-means, PCA, Association Rule Learning, Casual Impact Analysis, Neural Networks, Deep Learning
- Skills: Applied Machine Learning and data analysis skills to get actionable insights from both structured and unstructured data to support stakeholders' decision making, MS Access database design and development. QA analysis (creation and implementation of test plans, scenarios, scripts. QA gap analysis for UAT and Production). Collaborating with team members.

DEGREE AND LICENSE

Master of Applied Data Science in Python Online , University of Michigan, Ann Arbor	Jan 2020 - Sep 22, 2022
Awarded MADS Pioneer Scholarship. Achieved a cumulative 3.974 GPA on a 4.0 scale.	
CPA (Chartered Professional Accountant), CPA Ontario, Toronto	2014
CMA (Certified Management Accountant), CMA Ontario, Toronto	2008
Postgraduate Diploma of Applied Information Technology , ITI Information Technology Institute, Ottawa, Canada	2001
Master of Business Administration (MBA), University of Ottawa, Canada	2000
Provincial Life License Qualification Program (LLQP) , Completed and Passed, Expired	May 8, 2015
Canadian Investment Funds Course (CIFIC) , Completed and Passed, Expired	Feb 27, 2015
Certificate of Accomplishment – Income Tax Preparation Level 2, H&R Block	Dec 10, 2011

CERTIFICATES

Certificate of Managing Project Risks and Changes , Coursera	Sep 2024
Certificate of Project Risk Management , Coursera	Sep 2024
Certificate of Decision-Making and Scenarios , University of Pennsylvania (Wharton) on Coursera	Sep 2024
Certificate of Regression Analysis Simplify Complex Data Relationships , Google on Coursera	Sep 2024
Certificate of Modeling Risk and Realities , University of Pennsylvania (Wharton) on Coursera	Sep 2024
Certificate of Introduction to Microsoft Power Platform , Microsoft on Coursera	Sep 2024
Certificate of Introduction to Spreadsheets and Models , University of Pennsylvania (Wharton) on Coursera	Sep 2024
Certificate of Fundamentals of Quantitative Modeling , University of Pennsylvania (Wharton) on Coursera	Sep 2024
Certificate of The Power of Statistics , Google on Coursera	Sep 2024
Project Certificate of Automating your BigQuery Data Pipeline with Cloud Dataprep , Google Cloud on Coursera	Aug 2024
Certificate of Microsoft Azure SQL , Microsoft on Coursera	Aug 2024
Project Certificate of Preparing and Aggregating Data for Visualizations using Google Cloud Dataprep , Coursera	Aug 2024
Project Certificate of Working with Cloud Dataprep on Google Cloud , Google Cloud on Coursera	Aug 2024
Certificate of BigQuery for Data Analysts , Google Cloud on Coursera	Aug 2024
Certificate of Microsoft Power BI Data Analyst , Microsoft on Coursera	Jul 2024
Project Certificate of Data Cleaning in Snowflake: Techniques to Clean Messy Data , Coursera	Mar 2024
Certificate of Programming Foundations with JavaScript, HTML, and CSS , Duke University on Coursera	Jan 2023
Certificate of Excel Power Tools (Power Query, Power BI) for Data Analysis , Macquarie University on Coursera	2022
Certificate of Creating Dashboards and Storytelling with Tableau , University of California, Davis on Coursera	2019
Certificate of Managing Big Data in Clusters and Cloud Storage , issued by Cloudera on Coursera	2019
Certificate of Statistics with SAS , issued by SAS Canada on Coursera	2019
Certificate of Java Essentials for Hadoop , Simplilearn	Dec 2016
Certificate of Introduction to MongoDB , issued by MongoDB Inc.	2019
SPSS (Statistical Package for Social Sciences), OptionTrain, Toronto, Completed	2017

Agnes Yunhong He

Data Science and Data Analytics

Yunhong@umich.edu cell: 647-608-3939 618-99 Cameron Street, Toronto, ON, Canada
<https://www.linkedin.com/in/agneshe/>

SAS Basic and Advanced Programming, Metro College of Technology, Toronto, Completed

2012

PROJECTS

- 1. Behavior-Based Customer Segmentation with K-Means, Data Science Infinity, England** **Oct 1, 2025 – Jan 2026**
 - Applied k-means clustering on grocery transaction data to divide the customers into distinct groups based on their shopping behavior, which could be used to better understand customers over time, more accurately target customers with relevant content & promotions, and save marketing costs.
- 2. Market Basket Analysis Using Association Rule Learning, Data Science Infinity, England** **Oct 1, 2025 – Jan 2026**
 - Used unsupervised association rule learning on grocery transaction data to identify frequently co-purchased products and support market basket analysis and bundled promotion strategies to boost sales.
- 3. Album Purchase Prediction with PCA Dimensionality Reduction, Data Science Infinity, England** **Oct 1, 2025 – Jan 2026**
 - Applied PCA for dimensionality reduction on customers' historical listening behavior across 100 artists' songs to predict album purchase decisions, increase model efficiency and reduce model computation costs.
- 4. Predicting Membership Sign-Ups Using Logistic Regression, Data Science Infinity, England** **Oct 1, 2025 – Jan 2026**

Used logistic regression to predict delivery club membership sign-ups based on customers' historical grocery shopping behavior, enabling targeted mailings to customers most likely to sign up instead of mailing everyone, significantly reducing mailing costs.
- 5. Capstone Project: Sentiment Analysis of Online Customer Reviews, University of Michigan** **Jun 28, 2022 – Aug 22, 2022**
 - Led sentiment analysis of Vietnamese e-commerce reviews using BERT and ML classifiers to identify mismatches between ratings and emotions. Built automated NLP pipeline and interactive dashboards; results supported team tasks.
 - Maintained public GitHub repository with code, notebooks, and visualizations; collaborated via Slack and Google Sheets. https://github.com/yunhonghe/realtime_dreamer.
 - Published sentiment analysis blog. <https://madsrealtimedreamer.wordpress.com/2022/08/18/task-2-sentiment-analysis/>
- 6. Predicting Text Difficulties, School of Information, University of Michigan** **Jul 27, 2021 – Sep 26, 2021**
 - Classified Wikipedia text as easy or difficult to read to help editors improve readability for readers with comprehension challenges, using **NLP and ML**, applied **PCA/SVD** for feature extraction, and documented results in a final report.
- 7. COVID-19 Pandemic Policy Analysis, School of Information, University of Michigan** **Mar 31, 2021 – May 31, 2021**
 - Collected and processed COVID-19 datasets, creating reusable Python functions for data cleaning, merging, and visualization; generated time series dashboards for cases, hospitalizations, vaccinations, and government responses in Canada and Ontario nursing homes.
 - Visualized global comparisons across the top 10 countries by vaccination, GDP, demographics, and health factors; analyzed COVID-19 impacts in Ontario nursing homes, India, Canada, and the USA.
 - Produced final reports and visualizations for all levels of Canadian government, highlighting the impact of government responses on infections, deaths, and recoveries to support evidence-based policy improvements.
- 8. MCQ (Multiple Choice Question) Grader, School of Information, University of Michigan** **Aug 19, 2020 – Nov 21, 2020**
 - Evaluated and graded the quality of the MCQ created by the students using Regex expressions, NumPy series and functions, Pandas data frames and functions in Python in the MOOC "Data Manipulation in Python" in the experiment designed by a PhD student supervised by the Professor for the course.

WORK EXPERIENCE

- Python Coding Expertise for AI Training, Outlier AI, USA, remotely working from home** **May 2024 – Now**
- Worked as a contributor/attempter to help Outlier AI train LLMs (Gen AI) to outperform existing ones by pushing the boundaries of what LLMs can do and enhancing their ability to process and generate human-like language. This enables their customers to create advanced AI applications that are effective and reliable.
 - Wrote creative Python code-related prompts and generated responses to beat the best AI model (SOTA) that would provide.
 - Completed rating assessments across the various dimensions for the two human-like responses generated by LLMs.

QA Lead/Senior Data Analyst, CITI via HCL Technologies, Canada

Mar 2022 – Dec 2022

Agnes Yunhong He

Data Science and Data Analytics

Yunhong@umich.edu cell: 647-608-3939 618-99 Cameron Street, Toronto, ON, Canada
<https://www.linkedin.com/in/agneshe/>

- Collected business requirements, performed QA, data validation, and gap analysis for banking market data onboarding.
- Built data models, ETL pipelines, and dashboards (Power BI, Tableau) to identify patterns and anomalies.
- Processed big data with PySpark, SQL, and Python, transforming unstructured data; documented defects in Jira and analyses in Confluence.

Senior Data Analyst, Sinai Health System via Robert Half, Toronto

Feb 2021 – Oct 2021

- Researched healthcare KPIs and hospital benchmarking; provided actionable recommendations to help secure additional government funding.
- Built MS Access databases, Excel financial models, and Tableau dashboards to analyze hospital performance and identify anomalies, supporting improved clinical outcomes and increased government funding.
- Developed ARIMA forecasts in Python for clinical program expenses to enable data-driven budgeting and strategic planning.

Instructional Aide, School of Information, University of Michigan

Sep 4, 2020 – Oct 8, 2020

Provided feedback (without grades) on the manifesto and graded assignments for the course “Being a Data Scientist.” Collaborated with the team via Google Sheets, team meetings, and Slack.

Data Analyst/MS Access Database Developer, Legal Aid Ontario via Altis Professional, Toronto, Canada Apr 2019 – Jun 2019

Project 1: Automation of Court Information Requesting Process and Reporting

- Designed and developed MS Access ACIR system with ETL, Snowflake schema, and SQL, improving court data retrieval from 15–30 days to 1–2 days.
- Created process maps, test plans, and user manuals; implemented functions to detect suspicious user activity.

Project 2: Automation of Importing Process of Monthly Court Information Text Files

- Designed and developed an MS Access application with ETL, data mapping, and Snowflake schema modeling for monthly court data; created test plans, procedures, and user manual.

Senior Data Analyst, RBC via Procom, Toronto, Canada

Jun 2018 - Oct 2018

- Collected business requirements, performed ETL, data mapping, and modeling for RBC business customer data using complex SQL; automated reporting with SAS and conducted statistical analysis.
- Identified data irregularities and improved Canadian business client penetration rate by applying Python NLP, SAS Levenshtein matching, and Tableau visualizations to cleanse, mine, and analyze text data.

Senior QA Analyst, Deloitte via Eagle Professional, Toronto, Canada

Oct 2017 - Apr 2018

- Collected business requirements, translated them into SAS code, and automated QA processes; performed ETL, data mapping, and modeling, and created Excel dashboards to monitor data quality and patterns.
- Developed data quality test matrices and scripts for marketing database migration (SQL Server 2014 → 2016), merging data via complex SQL and performing validation and reconciliation.

Data Scientist/Senior QA Analyst, Telus via emergiTEL, Toronto, Canada

Jan 2017 – Jun 2017

- Managed 500TB of clickstream data, performing ETL, analysis (PySpark, Hive, SAS), and dashboards (Power BI, Tableau, Domo) to identify anomalies and provide actionable insights.
- Conducted A/B testing, KMeans clustering, and data quality controls, including validation, reconciliation, and automated reporting.

Data Analyst/Senior Risk Analyst, BMO Capital Markets via Procom, Toronto, Canada

Jun 2015 – Dec 2016

- Collected business requirements and designed MS Access applications with SQL/VBA to automate APL monitoring and reporting, performing ETL, data mapping, and integration from 15 trading systems; assessed product, currency, and market risks, helping the risk department significantly reduce trading risks and comply with government policies.
- Extracted data from IBM Netezza to Excel and prepared daily market risk limit reports for multiple portfolios.

Data Analyst, Canadian Imperial Bank of Canada (CIBC), Toronto, Canada

Dec 2012 – Nov 2014

- Collected business requirements and performed ETL, data mapping, and modeling for marketing campaign data; analyzed customer demographics, spending habits, and other characteristics using SAS/SQL/Macro.
- Conducted QA audits, data validation, and reconciliation; automated reporting processes using Unix.
- Calculated and visualized ROI and campaign performance, conducted A/B testing, and developed predictive models for Customer Lifetime Value (CLV).

Agnes Yunhong He

Data Science and Data Analytics

Yunhong@umich.edu cell: 647-608-3939 618-99 Cameron Street, Toronto, ON, Canada
<https://www.linkedin.com/in/agneshe/>

Data Developer/MS Access Application Developer, Altis Professional, Toronto, Canada

Oct 2011 – May 2012

Project 1: Automation of Forecast, Monitor and Reporting, Ministry of Training, Colleges and Universities, Mar 2012–May 2012

- Collected business requirements, addressed data quality and operational issues, and developed an MS Access Forecasting, Monitoring, and Reporting application using SQL/VBA with ETL, data mapping, and modeling, improving reporting efficiency.
- Created process flowcharts, user documentation, test plans, and delivered training and presentations; validated and reconciled data.

Project 2: Automation of Auditing Issues Tracking Reports, Ministry of Finance

Oct 2011 - Dec 2011

- Collected business requirements and developed an MS Access auditing issues tracking application with SQL/VBA, including ETL, data mapping, modeling, and Excel reporting.
- Created test plans, validated and reconciled data, and delivered visualizations, presentations, and user documentation.

Financial Analyst/ MS Access Application Developer, Bank of Montreal (BMO) via Robert Half

Mar 2011 – Jul 2011

Project: Automation of Management Reporting Process

- Collected business requirements and replaced a 20-year-old Excel reporting tool by designing and automating MS Access applications with SQL/VBA and Excel financial models, streamlining 22 monthly/quarterly management reports and reducing generation time from 6 hours to 0.5 hours.
- Created data mapping, modeling, and visualizations, identified abnormalities, assessed financial risks, and delivered user documentation.

Financial Analyst, FSS Management Reporting, TD Financial Group, Toronto, Canada

Jul 2009 – Jan 2010

- Collected business requirements, reviewed financial data, investigated abnormalities, and prepared management reports, using MS Excel (VBA), MS Access (SQL/VBA), PeopleSoft, Hyperion, INEA, and Crystal Reports.
- Designed financial models, performed variance analysis, identified risks, validated/reconciled data, and used SAS to extract and process transaction data.

Senior Financial Analyst, Aviva Canada via Inteqna, Toronto, Canada

Jul 2007 – Aug 2008

- Developed MS Access/Excel reporting applications with SQL/VBA for management and HR, performing ETL, reconciliation, and data modeling.
- Automated 130 PivotTables for actuals, forecasts, and budgets, reducing report prep time from 7 hours to 0.5 hours and highlighting financial risks.
- Delivered executive presentations and documented user manuals.

Financial Analyst, Canadian Imperial Bank of Canada (CIBC) via Manpower, Toronto, Canada

Jan 2004 – Jul 2007

- Investigated invoices and general ledger for data abnormalities, prepared credit card revenue, security lending, and credit risk reports, and automated reporting with VBA and SQL; conducted statistical analysis on cardholder demographics and spending.
- Managed AML and KYC processes, analyzing client accounts, investigating abnormalities, verifying signatories, and identifying risks/fraud using CIBC Siebel, COINS, and ADP.
- Analyzed collections and mortgage data in Excel, producing weekly reports recognized by senior management.