

Using Data Analytics within the Organization: A 'Low Hanging Fruits' Approach

Theo Stratopoulos, PhD



UNIVERSITY OF
WATERLOO

SCHOOL OF ACCOUNTING
AND FINANCE



Theo Stratopoulos

My teaching and research focus is on financial performance and competitive implications of IT-enabled business strategies and emerging technologies. Currently, I am working on projects related to emerging technology adoption, blockchain, and covid-19 related firm disclosures,. My research has been published in *The Accounting Review*, *Journal of Management Information Systems*, *Communications of the ACM*, *Journal of Information Systems*, *Information & Management*, *Journal of Strategic Information Systems*, and *International Journal of Accounting Information Systems*. I have consulted on data analytics projects and delivered workshops to professionals and academics on data analytics and emerging technologies. I have authored several open-source texts (available from [SSRN](#)) on data analytics, blockchain, and IT business value. I am a member of the CPA Canada - Audit Data Analytics committee, associate editor for the *International Journal of Accounting Information Systems*, and senior editor for the *Journal of Information Systems* of the American Accounting Association.

Presentation Focus

- Adopt a systematic approach to data analytics (CRISP-DM)
- Tool vs Concepts: Focus on Concepts
- Concepts to Invest on: Querying single/multiple data files & regression
- A demonstration

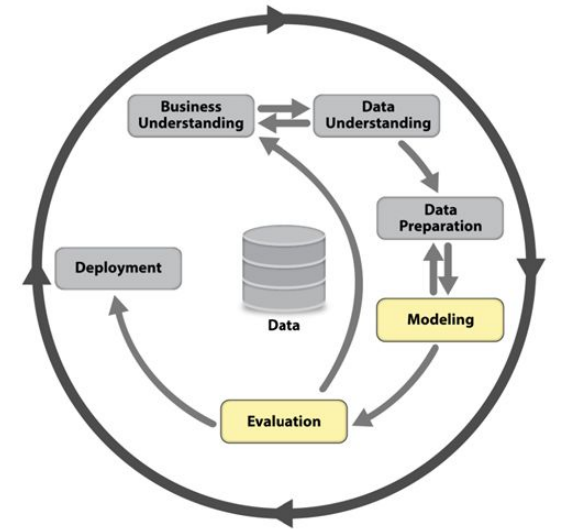
A Systematic Approach to Data Analytics (CRISP-DM)

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling & Evaluation
5. Communication



A Systematic Approach to Data Analytics: Why?

1. Business Understanding
 - **Focus on what matters**
2. Data Understanding
 - **Realistic Expectations/limitations**
3. Data Preparation
 - **Leverage your knowledge/contribute**
4. Modeling & Evaluation
 - **Focus on economic instead of statistical significance**
5. Communication
 - **Monetize your results - explore new business questions**



Tools vs Concepts:: Pick Concepts

- **What is the best data analytics tool?**
 - Excel
 - Power BI
 - Tableau Prep/Tableau
 - Alteryx
 - SAS
 - SPSS
 - R or Python

Tools vs Concepts:: Pick Concepts

- **What is the best data analytics tool?**
 - Excel
 - Power BI
 - Tableau Prep/Tableau
 - Alteryx
 - SAS
 - SPSS
 - R or Python
- **"It doesn't matter whether a cat is white or black, as long as it catches mice?"**

Concepts to Invest/Learn

- “The vast majority of accounting problems can be answered with well designed queries and basic statistics”

Concepts to Invest/Learn

- “The vast majority of accounting problems can be answered with well designed queries and basic statistics”
 - **Querying single data files**
 - **Querying multiple (relational) data files**
 - **Leverage regression to generate predictions**

Querying Single Data Files

From a data set (e.g., sales data)

Select variable (e.g., store, date, revenues)

Filter relevant rows (e.g., 1st quarter 2020)

Group by categories (e.g., group by store)

Aggregate values for target variables (e.g., sum of revenues, min/max)

Filter relevant aggregate values (e.g., top/bottom 5 stores based on revenues)

Querying Single Data Files

From a data set (e.g., sales data)

Select variable (e.g., store, date, revenues)

Filter relevant rows (e.g., 1st quarter 2020)

Group by categories (e.g., group by store)

Aggregate values for target variables (e.g., sum of revenues, min/max)

Filter relevant aggregate values (e.g., top/bottom 5 stores based on revenues)

- Apply this SQL logic to make better pivot tables!

Querying Multiple Data Files - Inner Join

<u>B&M Sales</u>		<u>Online Sales</u>		<u>Inner Join</u>		
ProductID	Store Price	ProductID	Online Price	ProductID	Store Price	Online Price
p001	\$9.99	p001	\$8.99	p001	\$9.99	\$8.99
p002	\$4.99					
		p003	\$5.50			
p004	\$12.45	p004	\$10.58	p004	\$12.45	\$10.58
p005	\$19.99					
p006	\$7.39	p006	\$7.02	p006	\$7.39	\$7.02

Querying Multiple Data Files - Left Join

B&M Sales

ProductID	Store Price	ProductID	Online Price
p001	\$9.99	p001	\$8.99
p002	\$4.99		
		p003	\$5.50
p004	\$12.45	p004	\$10.58
p005	\$19.99		
p006	\$7.39	p006	\$7.02

Online Sales

Left Join

ProductID	Store Price	Online Price
p001	\$9.99	\$8.99
p002	\$4.99	NA
p004	\$12.45	\$10.58
p005	\$19.99	NA
p006	\$7.39	\$7.02

Querying Multiple Data Files

What is the **kind of join** you need? (e.g., common fields for inner join vs all from one and matching from other for left join)

Select variable(s) **From** first table

Select variable(s) **From** second table

Identify **common field** in both tables to make the join

Continue with rest of steps as in querying single file

Querying Multiple Data Files

What is the **kind of join** you need? (e.g., common fields for inner join vs all from one and matching from other for left join)

Select variable(s) **From** first table

Select variable(s) **From** second table

Identify **common field** in both tables to make the join

Continue with rest of steps as in querying single file

- Apply this SQL logic to make better vlookups!

Leverage regression to generate predictions

“Whenever we create a graph of lets say sales against advertising or stock market returns over time, implicitly or explicitly we try to visualize the possible existence of a line going through our data points. Our attempt to add this line in our data, manifests our implicit or explicit assumptions that the future will look like the past or a new set of observations drawn from the same population would generate predictable results.”

Suggested Resources from my SSRN Page

1. Stratopoulos & Rogozynski “An Introduction to the Standard Data Analytics Process for Accounting Students”
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3512960
2. Stratopoulos and Shields “Basic Audit Data Analytics with R”
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3142207
3. Stratopoulos & Kennedy “Accounting and Finance Applications for Introduction to Statistics with R”
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3184054
4. Stratopoulos & Vanden Bosch “A First Year Data Analytics Course for Accounting Students”
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3618697

WATSPEED

[WatSPEED home](#)[About WatSPEED](#) >[News](#)[Programs](#) v[Professional development](#)[Executive programs](#)[Custom training](#)[Contact us](#)[INFORMATION FOR](#)[MyCreds at Waterloo](#)[WatSPEED » Programs »](#)

Digital Certificate in Predictive Analytics for CPAs



The banner features a dark blue background with a yellow diagonal stripe at the bottom. It includes the University of Waterloo School of Accounting and Finance logo, the CPA Ontario logo, and the text 'STRATEGIZE AND LEAD WITH DATA' in large white letters. Below this, a yellow box contains the text 'Turn insights into transformative decision making.' A yellow arrow points from the bottom right towards the center. The bottom right corner features the University of Waterloo logo and the text 'WAT SPEED'.

UNIVERSITY OF
WATERLOO
SCHOOL OF ACCOUNTING
AND FINANCE

CPA
ONTARIO

STRATEGIZE AND LEAD WITH DATA

Turn insights into transformative decision making.

UNIVERSITY OF
WATERLOO | **WAT SPEED**

[ENROL IN THIS COURSE VIA CPA ONTARIO](#)[Register now](#)[GET MORE DETAILS ABOUT THIS COURSE](#)[Download brochure](#)

Integrated Case Demonstration

Business Understanding

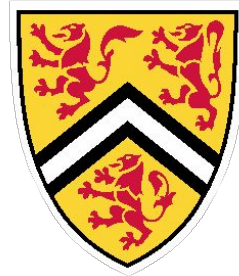
- A privately held business headquartered in Maine, prides itself on offering creative soups at a reasonable price and that are made with locally sourced ingredients.
- The company sells their soup at several restaurant locations throughout Maine and divided in geographic region: northern Maine (1), mid-Maine (2), and coastal Maine (3).
- An audit of the company is required to comply with debt covenants related to a large bank loan that the company entered into when it began operations.

Audit Problem

- As part of your team's role in this audit engagement, you will be performing a large portion of the planning and testing of sales for the 2016 audit.
- In prior years, the audit approach relied on random sampling to test revenues. However, your team wants to develop more focused procedures in the current year to hone in on potentially riskier sales transactions. As a result, the plan is to perform predictive sales analytics to identify unusual trends in the daily sales data with the goal of identifying sales on specific days at specific store locations that should be subjected to substantive testing due to heightened risks. The remainder of the population would then be sampled using a random sampling approach.

Demonstration with R

UNIVERSITY OF
WATERLOO



SCHOOL OF ACCOUNTING
AND FINANCE

Using Data Analytics within the Organization: A 'Low Hanging Fruits' Approach

Theo Stratopoulos, PhD