Data Analytics

The increasing amount of data in our world today has been both a benefit and detriment to the modern accountant. The sheer volume of data can be particularly useful in analyzing performance, solving problems, and producing evidence. At the same time, this abundance of data can create information overload as well as challenges to interpreting the data to make useful conclusions.

There are many types of data, such as numeric data, text data, dates, times, etc. Data needs to be stored somewhere, typically in files (such as an Excel of CSV file) or databases. Data can be structured, such as being bound to a table, or unstructured, like a text file output of social media posts. Most accounting systems are set up to handle structured data.

Data can also be created through transactions. For example, data can be created by a human inputting fields into a form, or it could be system generated, like a time stamp showing when a user accessed a particular file. When data is untampered, it is referred to as raw data. Raw data is the purest form of data, but it is often not useful until it is processed, or analyzed, by systems and humans. Processed data can be useful for understanding complex situations, but processed data can also contain human errors or biases. It is therefore important to understand how the data was processed.

I. How Different Entities Use Tax Data for Analysis

Government and individuals use data to achieve different goals.

Tax departments

Tax departments in large organizations use data to assist with tax compliance issues. This could include routine tasks like evaluating the tax liabilities of the entity, along with more complex analyses such as determining what would happen to the organization's effective tax rate if it were to set up a new permanent establishment in a new province/territory or foreign jurisdiction.

Broadly speaking, larger organizations often have a hierarchy in their organizational structure: An operations level typically handles day-to-day transactions, a middle-management level handles tactical decision-making, and an executive level oversees the organization's strategic objectives. Organizations often use different types of information systems to ensure that the right people are able to access relevant information at the appropriate time. Some examples of different information systems and their purpose are described below:

- Transaction processing systems (TPS) are designed to facilitate day-to-day operations, such as payroll, GST/HST reporting, processing accounts receivable and accounts payable, etc.
- Management information systems (MIS) serve the middle management level of the organization. They provide managers with reports about the organization's current performance or make comparisons with historical records. They are typically used for internal purposes, not external reporting, and serve the functions of planning, controlling, and decision-making at the management level.
- Decision support systems (DSS) help management make decisions that are more complex in nature. These systems often involve more use of data analytics to model and interpret data, or to condense large amounts of data in a useful form that can be analyzed by decision-makers.
- Executive support systems (ESS) help senior management and executives make strategic decisions for the organization. They help facilitate non-routine decision-making involving complex issues, judgment, and insight.

Professional advisors and tax preparers

Advisors often use data for tax planning. Historical data can be used to advise on changes in tax rules, corporate reorganizations, expansion into new markets, and transfer pricing for inter-company transactions. Data is also often used to quantify the tax exposure when

an uncertain tax position is taken. For example, a client may have a defensible position that their workers are independent contractors. However, as discussed in Chapter 4, determining if workers are contractors or employees can be a grey area. The advisor could therefore use data to analyze the amount of payroll tax and withholdings that would be owed if the CRA were to successfully argue that the workers are employees. Depending on the size of the exposure and the client's risk tolerance, this data could influence the decision on whether to keep these workers as independent contractors or to put them on the payroll as employees.

More data is also beneficial to the completion of tax returns, and tax compliance technologies have advanced to automate much of the tax preparation process. Most current tax software contain things like summaries of tax returns for the last five years. This can help the tax preparer identify trends to locate missing information. For example, suppose a client had significant dividend income for the past several years and the tax return drafted for the current year shows a significant drop in dividend income. This could be due to missing information, and the tax preparer can follow up with the client to determine if there has been a change in investment activity or if some income was just missed.

Canada Revenue Agency (CRA)

The CRA uses data analytics to detect non-compliance. Taxpayers would likely be surprised by the amount of information the CRA has access to. Many sources of income are reported to the CRA even before an individual taxpayer has a chance to file their tax return. By the time the April tax season rolls around, an individual will have already received various slips, such as:

- **T4 slip**—Discloses the amount of employment income, including taxable benefits, earned by an individual. Payroll deductions such as Canada Pension Plan (CPP) contributions, Employment Insurance (EI) premiums, and taxes are also reported on this slip.
- **T5 slip**—Discloses most sources of property income, such as interest, dividends, and royalties.
- **T3 slip**—Discloses any distribution made by a trust to its beneficiaries. Investors in commercial trusts (Chapter 17) will also receive these slips for any income they receive on their investment.
- **T5013 slip**—Discloses income earned by a partnership and allocated to its partners.
- **T4RSP slip**—Discloses any withdrawals from an RRSP.

The CRA applies data analytics to match up amounts disclosed on these slips to the actual tax returns filed. Taxpayers who do not report the amounts on these slips, whether intentionally or inadvertently, will eventually get either a letter or automatic reassessment reconciling for any missing slips. The CRA has become so proficient at this that these reassessments or letters can be issued to the taxpayer without any human intervention.

The CRA also receives personal and financial data from other sources. For example, information regarding title transfers of real property is disclosed to the CRA. This data could be used to identify real estate transactions reported as capital gains that are at a higher risk of being classified as business income, or to identify inappropriate claims of the principal residence exemption.

Banks and other financial institutions are required to disclose certain financial data to the CRA, often for anti-money laundering purposes. For example, electronic fund transfers (EFTs) of more than $10,000, or smaller transfers totalling this amount in a short period of time, are disclosed to CRA. Social media posts flaunting extravagant lifestyles can also be used in situations where the CRA is performing a net worth analysis to assess the risk of an individual having unreported income.

The CRA's mandate is to ensure that taxpayers are paying the correct amount of tax. By using data analytics, the CRA can identify the taxpayers who are at a higher risk of reporting an incorrect amount of tax.

Department of Finance

While the CRA is responsible for administering Canada's income tax laws, a common misconception is that the CRA also creates the law. Our tax laws are actually introduced by the Department of Finance, and data analytics can play a major role in developing new tax policy.

The Department of Finance may engage in *what-if scenarios* to estimate the impact of a proposed tax change. A what-if scenario involves the manipulation of inputs (e.g., tax rates, income thresholds, etc.) to estimate a future outcome. The goal with this type of analysis is to achieve a desired outcome by using the optimal inputs. Various input combinations are often considered, along with the expected outcome and the probability of the outcome occurring.

For example, several years ago there were sweeping changes to the taxation of Canadian-controlled private corporations. The small business deduction was increased, the refundable tax mechanics were modified, and adjusted aggregate investment income was introduced to reduce access to the small business deduction for corporations with too much passive income (topics discussed in Chapter 13). It is highly likely that what-if scenarios were used to determine the increase to the small business deduction and the appropriate reduction to the $500,000 business limit from adjusted aggregate investment income. In the end, the business limit changes resulted in a $1.283 billion tax increase over five years on a change that only affected 3% of corporations.[1]

II. Storing Data

When it comes to accounting and tax records, data is often stored in a *relational database*. Databases can store more data, perform faster, and provide more security; however, connecting to and extracting data from a database can be challenging. A relational database uses a structure that allows users to identify and access data in relation to another piece of data in the database. This data is often organized into two or more tables consisting of columns and rows. A relational database is typically more optimal than storing data in a flat file like a single Excel document since it ensures the data is complete, reduces redundancy in the data, and allows for better internal controls to be implemented over the access, use, and safeguarding of the data.

Every column in a table should be unique and serve a purpose. Columns in a table can generally be classified as follows:
- **Primary keys—**Identifiers used to ensure that each row in the table is unique. For example, a table consisting of information on various taxpayers could use their Social Insurance Number as the primary key. Another example could be the employee identification number in a table consisting of sensitive payroll information.
- **Foreign keys—**Identifiers that reoccur within a data set.

To illustrate the relationship between primary (PK) and foreign keys (FK), consider a very simple database that consists of three tables:
- Table 1 contains a list of products sold by the entity with unique product ID numbers (PK).
- Table 2 contains a list of customer names (PK).
- Table 3 contains sales information including invoice numbers (PK), the product sold (FK), and the customer the product was sold to (FK).

These tables form a relationship. In Table 3, the same product or customer could appear multiple times in the data set. Meanwhile, the invoice number is used to ensure each sale is uniquely maintained. The foreign keys in Table 3 are primary keys in Tables 1 and 2, which

---

1   https://www.canada.ca/en/department-finance/news/2018/04/backgrounder-tax-fairness-for-the-middle-class-and-opportunity-for-all-canadians.html#ftn1 and https://www.budget.gc.ca/2018/docs/tm-mf/si-rs-en.html#Toc507170848

is where the data regarding the list of products and customer information is maintained. This data only appears once in their respective tables.

A full discussion on how databases are used to maintain accounting and tax records is beyond the scope of this text.

III. Exporting Data

Data may need to be extracted from a database for various tax-related reasons. For example, consider a condo developer that incorporates a new company to begin a new construction project. The new corporation files a GST/HST return claiming a substantial refund. To ensure the integrity of the GST/HST system, the CRA selects the return for a review and requests a listing of the input tax credits claimed. There are too many transactions to manually list out, so the data is extracted from the accounting system.

Several procedures should be followed when data is extracted from the accounting system:

1. **Determine the purpose and scope of the data extraction—**In this example, the purpose of the extraction is to respond to a CRA request for information. The scope should be limited to the reporting period under review.

2. **Obtain the data—**An organization with proper segregation of duties will likely not allow full access to the accounting database to a single person. An individual may have the authorization they need to access the data required to respond to CRA, or they may need to make a request for this information. Each organization can have different procedures in place for how to properly request the data and/or authorize its release to the CRA.

3. **Validate the data—**Whenever data is moved from one place to another, it is possible that data could have been lost or corrupted during the transfer. Data should be validated for completeness and integrity. For example, the number of records exported from the system could be compared to the number of records in the source database.

4. **Clean the data—**When data is exported from the database, formatting inconsistencies could occur (e.g., provinces may be abbreviated differently in different tables, or negative numbers may be presented in brackets in some places while a minus sign is used in other places). Unnecessary headings or subtotals or other issues may also occur. The data may need to be cleaned up to create a format that makes it more accessible to its user and for its purpose. Care needs to be taken at this step to ensure that the data is not altered when it is cleaned up.

After these procedures, the data is generally ready for analysis. In our example with the condo developer, the analysis would be done by the CRA.

IV. Analyzing Data

Data can be analyzed in numerous ways. Broadly speaking, there are four types of model-building methodologies:

- *Descriptive analytics* use historical data or events to better understand changes that have occurred in a business.
- *Predictive analytics* use statistical techniques or algorithms to determine the probability of an expected outcome or event.
- *Diagnostic analytics* examine data to understand why a particular thing is happening. They could also be used to troubleshoot an issue. For example, most tax preparation

software includes diagnostic tools that determine what inputs into the tax return are preventing the return from being electronically filed.

- *Prescriptive analytics* use data to determine the best course of action in a given situation.

Many tools are available to analyze data, depending on the purpose and scope of the analysis. For example:

1. *Data manipulation tools* are used for tasks such as performing calculations, filtering, and pivoting data. Excel and Alteryx are examples of software typically used for this purpose.

2. *Data visualization tools*, as the name suggests, help display and interpret data visually, such as through interactive charts and graphs. Power BI and Tableau are commonly used for this purpose.

3. *Audit-specific data analytic tools* help transform financial data into actionable audit risk information. These tools allow audits to be completed with less manual work while providing a higher level of assurance. An example of this tool would be Caseware IDEA.

The remainder of this Appendix will consist of a tutorial to complete some basic data analysis in a routine tax situation. It is intended for beginners; advanced techniques are beyond the scope of this text.

Although the CPA program uses a type of software called Power BI to expose candidates to data analytics, we have opted to use Microsoft Excel for this tutorial because it is readily available.

TUTORIAL DUNCAN'S COAT EMPORIUM

Situation: Duncan's Coat Emporium (Duncan's) is wholesaler of affordable outerwear. The company ships products to retailers all over Canada. The company reported that it collected $5,543.43 of GST/HST (Line 105) in the 2022-01-01 to 2022-03-31 reporting period. You are the company's controller and you want to analyze whether the company charged the correct GST/HST based on the sales made during this period. You have exported some sales data from the company's accounting system and you have already validated the export for completeness and integrity.

Required dataset: Open the Buckwold Tutorial Excel file available on Connect under Student Resources.

Purpose and scope: The goal is to analyze the amount of GST/HST payable on sales in the 2022-01-01 to 2022-03-31 reporting period. How can you determine the appropriate amount of GST/HST collected without manually isolating the transactions in this date range and without manually computing the GST/HST on each transaction?

Analysis: Columns A to F in the dataset contain information about the sales made to various customers. Columns K to L contain GST/HST information for the provinces and territories.

- **Invoice_ID**—Unique identifier for each sales invoice
- **Order_Date**—Date on which the order was placed.
- **Quantity_Sold**—Quantity of product sold on the particular transaction

- **Product_Description**—Description of the item sold
- **Product_Price**—Price of each product sold in the transaction
- **Store_Location**—The province or territory in which the customer is located
- **Prov**—The provinces and territories in abbreviated format
- **GST_HST_Rate**—The tax rate for the province or territory

Step 1: Formatting the Data into Tables

First, format the data into tables. This will reduce a lot of manual processes later when the data is analyzed.

(a) Select all the cells from A1 to F157. A quick way to do this without using your mouse is to hold SHIFT+CTRL and press DOWN and RIGHT on your keyboard's arrow keys.

(b) With the cells selected, press CTRL+T to create a new table. You may be prompted to confirm the data you have selected (press OK).
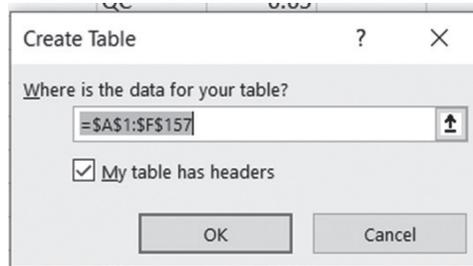
(c) You can change the name of your table in the "Table Design" tab that should now be visible at the top of your screen. Rename the table from "Table1" to "sales_transactions."
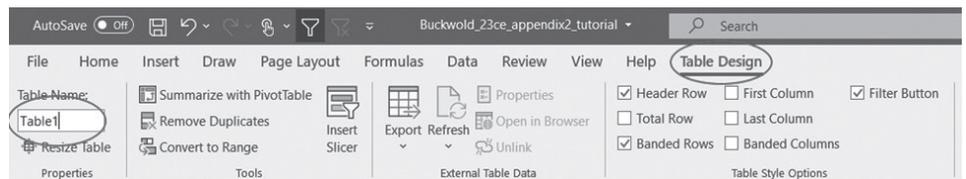
(d) Repeat the steps above to create a second table from cells K1 to L14. You can return to the top of your excel sheet without using your mouse by pressing CTRL+UP. Name your second table "province_rates." The result should look like this:

Prov	GST_HST_Rate
NL	0.15
PE	0.15
NS	0.15
NB	0.15
QC	0.05
ON	0.13
MB	0.05
SK	0.05
AB	0.05
BC	0.05
YT	0.05
NT	0.05
NU	0.05

Step 2: Using VLOOKUP

To calculate the amount of GST/HST owed for each transaction, match the province or territory the product was shipped to with the correct GST/HST rate. This would be tedious to complete manually, and doing so would introduce a higher risk of human error.

Use Excel's VLOOKUP function to easily match the province/territory where each sale was made with the appropriate tax rate.

(a) Add a new column to the sales_transactions table in column G, next to the Store_Location column. Name the column "Province_GST_HST_Rate."

(b) In Cell G2, apply VLOOKUP with the following parameters:

i. Cell_reference: This needs to reference the row's corresponding "Store_location" value. Excel will match this location with the corresponding province in the sales tax table.

ii. Table_array: This needs to reference the "province_rates" table.

iii. Column_number: This needs to reference the number of the column that contains the descriptive data that needs to be returned (i.e., the GST/HST rate is located in the second column of the "province_rates" table, so the number "2" should be used).

iv. True or false: Since you are looking for an exact match, this should be entered as FALSE.

v. Hit enter and column G should automatically populate.

Your screen should look something like this:

F	G	H	I	J	K	L
Store_Location	Province_GST_HST				Prov	GST_HST_Rate
BC	=VLOOKUP([@[Store_Location]],province_rates,2,FALSE)				NL	0.15
ON	VLOOKUP(**lookup_value**, table_array, col_index_num, [range_lookup])				PE	0.15
SK					NS	0.15
SK					NB	0.15
ON					QC	0.05
BC					ON	0.13
NL					MB	0.05
ON					SK	0.05
MB					AB	0.05
QC					BC	0.05
ON					YT	0.05
QC					NT	0.05
ON					NU	0.05

(c) Add a column to the sales_transactions table in column H and name it "GST_HST_Owing." In Cell H2, multiply "Quantity_Sold" by the "Product_price" and "Province_GST_HST_Rate." The entire column H should populate with this formula once you hit enter:

C	D	E	F	G	H	I	J	K
Quantity_Sold	Product_Description	Product_Price	Store_Location	Province_GST_HS	GST_HST_Owing			Prov	GST_HS
9	Cinched trench coat	107	BC		0.05	=[@[Quantity_Sold]]*[@[Product_Price]]*[@[Province_GST_HST_Rate]]		

The formula in H2 is   =[@[Quantity_Sold]]*[@[Product_Price]]*[@[Province_GST_HST_Rate]]

(d) You now have the GST/HST calculated for every transaction! The last thing to do is to determine how much of this GST/HST falls within the 2022-01-01 to 2022-03-31 reporting period.

Step 3: Using SUMIFS

(a) In any blank cell, apply SUMIFS with the following parameters:

i. Sum_range: This needs to reference the column that contains the data you wish to sum up only when specific criteria are met. This should reference the "GST_HST_Owing" column since you want to sum up the GST/HST in a specific date range.

ii. Criteria_range1: This needs to reference the range of data you want to consider for one of the conditions that will determine whether you pick up the GST/HST from a particular row. This should be set up to reference the "Order_Date" column, since you want to analyze if the "GST_HST_Owing" for a particular transaction falls into the date range.

iii. Criteria1: This can be used to set up a logical parameter. In order to set the earliest date in the range (January 1, 2022), input the criteria as ">=2022-01-01" (including quotation marks).

iv. Criteria_range2: This needs to reference the Order_Date column, since you want to use this data to set up the upper limit to the date search.

v. Criteria2: In order to set the latest date in the range (March 31, 2022), input the criteria as "<=2022-03-31" (including quotation marks).

Following these steps effectively creates an IF/THEN statement to pick up the GST/HST paid only if the transaction falls within the 2022-01-01 to 2022-03-31 date range.

B	C	D	E	F	G	H	I	J	K	L
Order_Date	Quantity_Sold	Product_Description	Product_Price	Store_Location	Province_GST_HST	GST_HST_Owing			Prov	GST_HST_Rat
2021-02-15	9	Cinched trench coat	107	BC	0.05	48.15	=SUMIFS(H:H,B:B,">=2021-01-01",B:B,"<=2021-03-31")			
2021-02-15	11	Cinched trench coat	107	ON	0.13	153.01	SUMIFS(sum_range, **criteria_range1**, criteria1, [criteria_range			

The formula in I2 is =SUMIFS(H:H,B:B,">=2022-01-01",B:B,"<=2022-03-31")

Conclusion

The result of this analysis is that the total GST/HST that should have been collected in the 2022-01-01 to 2022-03-31 reporting is $6,652.77. Since this is more than the actual GST/HST reported as collected in line 105 of the return, you have determined that Duncan's has remitted an incorrect amount of tax.

To correct this issue, the 2022-01-01 to 2022-03-31 reporting period can be amended by accessing the company's "My Business Account" on the CRA website. The balance due date for this return is 2022-04-30. Any increase to the GST owing after this date will be subject to interest at a CRA prescribed rate (currently a 5% annual rate compounded daily).

In situations where a very large amount of unremitted GST is uncovered, it may be beneficial to consider the CRA's Voluntary Disclosure Program (VDP) to rectify the situation. The VDP program allows taxpayers and GST registrants the opportunity to come forward with corrections while potentially receiving a reduction to interest and penalties. Relief is granted at the CRA's discretion based on the facts of the particular situation, but interest relief is usually limited to 50%. It would not be necessary for Duncan's to use a VDP for this correction since the error is relatively small and the potential benefit (interest savings) would need to be weighed against the costs (e.g., professional fees) to apply under the VDP.

Test your knowledge

Your instructor may ask you to apply what you have learned by answering the following questions using the tools showcased in this tutorial. For each question, include the amount as well as the formula you used to determine the answer.

1. What was the total GST/HST collected during the 2022-04-01 to 2022-06-30 reporting period?

2. What was the total amount of HST collected from provinces with a 15% HST rate?

3. How many parkas were sold in Ontario during December 2021?

4. How many products were sold for more than $100 during 2022?

Closing Remarks

This appendix provided a brief overview of data analytics in the context of Section 6.1.2. and 6.1.3. of the 2020 CPA Competency Map. For more detailed information about how data analytics are being applied to other CPA technical competencies, check out Section 7-7 of the competency map (available on the CPA Canada website). McGraw Hill also has several dedicated texts on data analytics, such as *Data Analytics for Accounting*, by Richardson, Teeter, and Terrell.