PCard-Expenditures-Analysis

August 20, 2023

1 Analysis of Toronto City Employee's Expense Card Transactions

1.1 Introduction

The City of Toronto is transparent about its employee expense card transactions, producing monthly reports available to the public. These transactions are accessible through the City's Open Data Portal.

1.2 Objectives

The core aims of this analysis include:

1. **Expenditure Characterization**: Understand the expenditure patterns of the City's Divisions or Cost Centres by exploring metrics like the number and frequency of transactions, their usual amounts, and their dispersion.

2. **Group Identification**: Determine significant groups of Divisions or Cost Centres with similar expense behaviours.

3. **Anomaly Spotting**: Uncover anomalies in the data that warrant further scrutiny.

1.3 Procedures

To accomplish these objectives, the following steps will be undertaken:

1.3.1 1. Data Collection and Preprocessing

1.1 Download the data from Toronto's Open Data Portal.

[ ]:

```
#1.1.1 Import required packages

from io import BytesIO
import matplotlib as mpl
import matplotlib.pyplot as plt

import os
import numpy as np
import pandas as pd
import re
import requests
import seaborn as sns
import zipfile

mpl.rcParams['figure.dpi'] = 300

# Shows plots in jupyter notebook
%matplotlib inline

# Set plot style
sns.set(color_codes=True)

import warnings

# Suppress the specific warnings
warnings.filterwarnings("ignore")
```

[ ]:

```
#1.1.2
# We need to load th PCard Expenditure data into one dataframe so that we can␣
 ↪work with them in Python.
# We have used API for accessing this data. Consuming an API is more efficient␣
 ↪and less prone to errors and thus preferred over other ways of loading data.

base_url = "https://ckan0.cf.opendata.inter.prod-toronto.ca"
url = base_url + "/api/3/action/package_show"
params = {"id": "pcard-expenditures"}

package = requests.get(url, params=params).json()

all_dataframes = []  # List to hold individual DataFrames

# Folder in the same directory as the Jupyter notebook
download_folder = "./"

for idx, resource in enumerate(package["result"]["resources"]):
    if not resource["datastore_active"]:
        resource_url = base_url + "/api/3/action/resource_show?id=" +␣
 ↪resource["id"]
        resource_metadata = requests.get(resource_url).json()
        download_url = resource_metadata["result"]["url"]
        resource_format = resource["format"].lower()
        filename = download_url.split('/')[-1]  # Extract filename from the URL

        print(f"Resource {idx} - URL: {download_url} - Format:␣
 ↪{resource_format}")

# Check if the resource format is zip
if resource_format == "zip":
    response = requests.get(download_url)
    with zipfile.ZipFile(BytesIO(response.content)) as z:
        # Count the number of .xlsx files within the zip
        xlsx_files_count = sum(1 for file in z.namelist() if file.endswith('.
 ↪xlsx'))

        for file in z.namelist():
            # Only load xlsx files within the zip
            if file.endswith('xlsx'):
                try:
                    df = pd.read_excel(z.open(file), engine='openpyxl')
                    all_dataframes.append(df)
                except Exception as e:
                    print(f"Error processing file {file} in {download_url}:␣
 ↪{e}")

        # Check if the number of .xlsx files matches the length of␣
 ↪all_dataframes
        if xlsx_files_count != len(all_dataframes):
            print(f"Warning: Number of xlsx files in the zip␣
 ↪({xlsx_files_count}) doesn't match the number of DataFrames loaded␣
 ↪({len(all_dataframes)}).")

        else:
            # Download and save (overwrite if exists) any other file
            response = requests.get(download_url)
            with open(os.path.join(download_folder, filename), 'wb') as file:
                file.write(response.content)
```

```
Resource 0 - URL: https://ckan0.cf.opendata.inter.prod-toronto.ca/dataset/ebc3f9
c2-2f80-4405-bf4f-5fb309581485/resource/070bdbd3-9bae-4269-b096-
e3a8bd7460c8/download/pcard_expenditures_readme.xls - Format: xls
Resource 1 - URL: https://ckan0.cf.opendata.inter.prod-toronto.ca/dataset/ebc3f9
c2-2f80-4405-bf4f-5fb309581485/resource/d83a5249-fb07-4c38-9145-
9e12a32ce1d4/download/expenditures.zip - Format: zip
```

1.2 Clean and consolidate the data into a single dataframe (combined_pcard)

[ ]:

```
#1.2.1
# Make a copy of all_dataframes that will not affect the original loaded␣
 ↪dataframe list.
pcard_dataframe_list = [df.copy() for df in all_dataframes]

# Get columns of the first dataframe as the reference column for consolidation␣
 ↪into a single dataset.
reference_columns = pcard_dataframe_list[0].columns

# Flag to indicate if structures are the same
```

```python
all_same_structure = True

# Iterate through all DataFrames in the list and  Check whether the dataframes in `all_dataframes` list have the same structure and column names
for idx, df in enumerate(pcard_dataframe_list[1:], start=1):  # Start from the second dataframe
    # If columns of current dataframe do not match reference columns
    if not df.columns.equals(reference_columns):
        print(f"DataFrame at index {idx} does not have the same structure as the first DataFrame.")
        all_same_structure = False
```

```
DataFrame at index 3 does not have the same structure as the first DataFrame.
DataFrame at index 5 does not have the same structure as the first DataFrame.
DataFrame at index 6 does not have the same structure as the first DataFrame.
DataFrame at index 7 does not have the same structure as the first DataFrame.
DataFrame at index 8 does not have the same structure as the first DataFrame.
DataFrame at index 9 does not have the same structure as the first DataFrame.
DataFrame at index 10 does not have the same structure as the first DataFrame.
DataFrame at index 11 does not have the same structure as the first DataFrame.
DataFrame at index 13 does not have the same structure as the first DataFrame.
DataFrame at index 14 does not have the same structure as the first DataFrame.
DataFrame at index 15 does not have the same structure as the first DataFrame.
DataFrame at index 16 does not have the same structure as the first DataFrame.
DataFrame at index 17 does not have the same structure as the first DataFrame.
DataFrame at index 18 does not have the same structure as the first DataFrame.
DataFrame at index 19 does not have the same structure as the first DataFrame.
DataFrame at index 20 does not have the same structure as the first DataFrame.
DataFrame at index 21 does not have the same structure as the first DataFrame.
DataFrame at index 22 does not have the same structure as the first DataFrame.
DataFrame at index 23 does not have the same structure as the first DataFrame.
DataFrame at index 24 does not have the same structure as the first DataFrame.
DataFrame at index 26 does not have the same structure as the first DataFrame.
DataFrame at index 27 does not have the same structure as the first DataFrame.
DataFrame at index 28 does not have the same structure as the first DataFrame.
DataFrame at index 30 does not have the same structure as the first DataFrame.
DataFrame at index 31 does not have the same structure as the first DataFrame.
DataFrame at index 32 does not have the same structure as the first DataFrame.
DataFrame at index 33 does not have the same structure as the first DataFrame.
DataFrame at index 34 does not have the same structure as the first DataFrame.
DataFrame at index 35 does not have the same structure as the first DataFrame.
DataFrame at index 36 does not have the same structure as the first DataFrame.
DataFrame at index 37 does not have the same structure as the first DataFrame.
DataFrame at index 38 does not have the same structure as the first DataFrame.
DataFrame at index 39 does not have the same structure as the first DataFrame.
DataFrame at index 40 does not have the same structure as the first DataFrame.
DataFrame at index 41 does not have the same structure as the first DataFrame.
```

```
[ ]: #1.2.2
     #check which columns in the other dataframes that differ from the first␣
      ↪dataframe(reference_columns) in the `all_dataframes` list.

     # Get columns of the first dataframe as the reference
     reference_columns = set(pcard_dataframe_list[0].columns)

     # Print the reference columns
     print("Reference columns:", ', '.join(reference_columns))
     print("")

     # Flag to indicate if structures are the same
     all_same_structure = True

     # Iterate through all DataFrames in the list
     for idx, df in enumerate(pcard_dataframe_list[1:], start=1):  # Start from the␣
      ↪second dataframe
         current_columns = set(df.columns)

         # Find columns that are in the reference but not in the current dataframe
         missing_in_current = reference_columns - current_columns
         # Find columns that are in the current dataframe but not in the reference
         additional_in_current = current_columns - reference_columns

         if missing_in_current or additional_in_current:
             all_same_structure = False
             print(f"DataFrame at index {idx} differs from the first DataFrame:")

             if missing_in_current:
                 print(f"  Missing columns: {', '.join(missing_in_current)}")
             if additional_in_current:
                 print(f"  Additional columns: {', '.join(additional_in_current)}")
             print("")
```

```
Reference columns: Merchant Type, Division, Cost Centre / WBS Element / Order
No., Batch Transaction ID, Transaction Date, G/L Account, Transaction Amt.,
Original Amount, Merchant Name, Trx Currency, G/L Account Description, Original
Currency, Purpose, Merchant Type Description, Card Posting Dt, Cost Centre / WBS
Element / Order No. Description

DataFrame at index 3 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No.
  Additional columns: Cost Centre / WBS Element / Order No

DataFrame at index 5 differs from the first DataFrame:
  Missing columns: Trx Currency
  Additional columns: Tr Currency
```

```
DataFrame at index 6 differs from the first DataFrame:
  Missing columns: Division, Cost Centre / WBS Element / Order No., Cost Centre
/ WBS Element / Order No. Description
  Additional columns: Divison, Cost Centre / WBS Element / Order, Cost Centre /
WBS Element / Order Description

DataFrame at index 7 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre /  WBS Element / Order No., Cost Centre /  WBS
Element / Order No. Decription

DataFrame at index 8 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS element / Order Description, Cost Centre
/ WBS element / Order

DataFrame at index 9 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description, G/L Account Description
  Additional columns: Cost Centre / WBS Element / Order , Cost Centre / WBS
Element / Order Description, Exp Type Desc

DataFrame at index 10 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description, G/L Account Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description, Exp Type Desc

DataFrame at index 11 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 13 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 14 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description
```

```
DataFrame at index 15 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 16 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 17 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 18 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 19 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description

DataFrame at index 20 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 21 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 22 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description, G/L Account Description
  Additional columns: Cost Centre/ WBS Element / Order, Cost Centre/ WBS Element
/ Order Description, Exp Type Desc
```

```
DataFrame at index 23 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 24 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 26 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 27 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 28 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Centre / WBS Element / Order, Cost Centre / WBS
Element / Order Description

DataFrame at index 30 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description, G/L Account Description
  Additional columns: Cost Center / WBS Element / Order, Exp Type Desc, Cost
Center / WBS Element / Order Description

DataFrame at index 31 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description

DataFrame at index 32 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description
```

```
DataFrame at index 33 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Batch Transaction ID,
Cost Centre / WBS Element / Order No. Description
  Additional columns: Cost Centre /WBS Element / Order, Cost Centre /WBS Element
/ Order Description, Batch-Transaction ID

DataFrame at index 34 differs from the first DataFrame:
  Missing columns: Division, Cost Centre / WBS Element / Order No., Cost Centre
/ WBS Element / Order No. Description
  Additional columns: Division , Cost Center / WBS Element / Order, Cost Center
/ WBS Element / Order Description

DataFrame at index 35 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description

DataFrame at index 36 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description

DataFrame at index 37 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBLS Element / Order Description, Cost
Center / WBS Element / Order

DataFrame at index 38 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description

DataFrame at index 39 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order #, Cost Center / WBS
Element / Order # Description

DataFrame at index 40 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description
```

```
DataFrame at index 41 differs from the first DataFrame:
  Missing columns: Cost Centre / WBS Element / Order No., Cost Centre / WBS
Element / Order No. Description
  Additional columns: Cost Center / WBS Element / Order, Cost Center / WBS
Element / Order Description
```

[ ]:

```python
#1.2.3
#By examining the output in 1.2.2 above it looks like the differences are in the column names. Some have spaces, others have fullstop etc.
#Thus we need to normalize the column names and use consistent names through out all the dataframes so that we can merge the data into one combined dataframe.

# `pcard_dataframe_list` is a list of your dataframes
reference_columns = pcard_dataframe_list[0].columns.tolist()

def normalize_column_name(col_name):
    col_name = col_name.lower()  # Convert to lowercase
    col_name = re.sub(r'[^a-z0-9]', '', col_name)  # Remove non-alphanumeric characters
    col_name = col_name.strip()  # Remove leading and trailing spaces
    return col_name

column_mapping = {
    'glaccountdescription': 'gl_account_description',
    'purpose': 'purpose',
    'cardpostingdt': 'card_posting_dt',
    'batchtransactionid': 'batch_transaction_id',
    'trxcurrency': 'trx_currency',
    'trcurrency': 'trx_currency',
    'originalcurrency': 'original_currency',
    'costcentrewbselementordernodescription': 'cost_centre_wbselement_ordernodescription',
    'merchantname': 'merchant_name',
    'transactionamt': 'transaction_amt',
    'costcentrewbselementorderno': 'cost_centre_wbselement_orderno',
    'transactiondate': 'transaction_date',
    'originalamount': 'original_amount',
    'glaccount': 'gl_account',
    'merchanttypedescription': 'merchant_type_description',
    'division': 'division',
    'divison': 'division',
    'merchanttype': 'merchant_type',
    'exptypedesc': 'gl_account_description',
    'costcentrewbselementorder': 'cost_centre_wbselement_orderno',
    'costcentrewbselementorderdescription':␣
 ↪'cost_centre_wbselement_ordernodescription',
    'costcentrewbselementordernodecription':␣
 ↪'cost_centre_wbselement_ordernodescription',
    'costcenterwbselementorder': 'cost_centre_wbselement_orderno',
    'costcenterwbselementorderdescription':␣
 ↪'cost_centre_wbselement_ordernodescription',
    'costcenterwblselementorderdescription':␣
 ↪'cost_centre_wbselement_ordernodescription',
    'costcenterwbselementorder#': 'cost_centre_wbselement_orderno',
    'costcenterwbselementorder#description':␣
 ↪'cost_centre_wbselement_ordernodescription',
}

# Normalize and map columns in each dataframe
for i, df in enumerate(pcard_dataframe_list):
    df.columns = [column_mapping.get(normalize_column_name(col), col) for col␣
 ↪in df.columns]

# Adjusting the reference dataframe's columns as well
reference_columns = [column_mapping.get(normalize_column_name(col), col) for␣
 ↪col in reference_columns]
pcard_dataframe_list[0].columns = reference_columns

# Combine all dataframes into one dataframe
combined_df= pd.concat(pcard_dataframe_list, ignore_index=True, sort=False)
```

```
[ ]: #1.2.4
     #Inspect the structure of the combined dataframe and see if everything checks␣
      ↪out.
     combined_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 220542 entries, 0 to 220541
Data columns (total 16 columns):
 #   Column                                     Non-Null Count   Dtype
---  ------                                     --------------   -----
 0   division                                   184088 non-null  object
 1   batch_transaction_id                       184088 non-null  object
 2   transaction_date                           184087 non-null  datetime64[ns]
 3   card_posting_dt                            184088 non-null  datetime64[ns]
 4   merchant_name                              184088 non-null  object
 5   transaction_amt                            218701 non-null  float64
 6   trx_currency                               218701 non-null  object
 7   original_amount                            220542 non-null  float64
 8   original_currency                          220542 non-null  object
 9   gl_account                                 184088 non-null  object
 10  gl_account_description                     184088 non-null  object
 11  cost_centre_wbselement_orderno             183456 non-null  object
 12  cost_centre_wbselement_ordernodescription  183847 non-null  object
 13  merchant_type                              184068 non-null  float64
 14  merchant_type_description                  184088 non-null  object
 15  purpose                                    182692 non-null  object
dtypes: datetime64[ns](2), float64(3), object(11)
memory usage: 26.9+ MB
```

[ ]:

```
#1.2.5
# Drop duplicate rows and reset index
combined_df = combined_df.drop_duplicates().reset_index(drop=True)
```

[ ]:

```
#1.2.6
#Check for missing/null values
combined_df.isnull().sum()
```

[ ]:

```
division                                     31042
batch_transaction_id                         31042
transaction_date                             31043
card_posting_dt                              31042
merchant_name                                31042
transaction_amt                               1316
trx_currency                                  1316
original_amount                                  0
original_currency                                0
gl_account                                   31042
gl_account_description                       31042
cost_centre_wbselement_orderno               31674
cost_centre_wbselement_ordernodescription    31283
merchant_type                                31062
merchant_type_description                    31042
purpose                                      32428
dtype: int64
```

[ ]:

```
#1.2.7
#Treat null values i.e. drop rows where division, batch_transaction_id,␣
 ↪merchant_type, card_posting_dt, merchant_name are null
#We are dropping the null values since our analysis and investigation relies␣
 ↪heavily on the batch ID and division hence dropping the rows without␣
 ↪batch_transation id
# will make the analysis more useful

mask = combined_df[['division','batch_transaction_id','transaction_date',␣
 ↪'merchant_type', 'card_posting_dt','merchant_name']].isnull().all(axis=1)
combined_pcard = combined_df[~mask]
combined_pcard.isnull().sum()
```

```
[ ]: division                                        0
     batch_transaction_id                            0
     transaction_date                                1
     card_posting_dt                                 0
     merchant_name                                   0
     transaction_amt                                 0
     trx_currency                                    0
     original_amount                                 0
     original_currency                               0
     gl_account                                      0
     gl_account_description                          0
     cost_centre_wbselement_orderno                632
     cost_centre_wbselement_ordernodescription     241
     merchant_type                                  20
     merchant_type_description                       0
     purpose                                      1386
     dtype: int64
```

```
[ ]: #1.2.8
     # Let us do further clean up by dropping observations with null transaction␣
      ↪dates.
     combined_pcard = combined_pcard.dropna(subset=['transaction_date'])
     combined_pcard.isnull().sum()
```

```
[ ]: division                                        0
     batch_transaction_id                            0
     transaction_date                                0
     card_posting_dt                                 0
     merchant_name                                   0
     transaction_amt                                 0
     trx_currency                                    0
     original_amount                                 0
     original_currency                               0
     gl_account                                      0
     gl_account_description                          0
     cost_centre_wbselement_orderno                632
     cost_centre_wbselement_ordernodescription     241
     merchant_type                                  20
     merchant_type_description                       0
     purpose                                      1386
     dtype: int64
```

```
[ ]: #1.2.9
     #Let us Inspect the strucutre and shape of our cleaned dataframe
     #Note: we have not treated the null values for purpose,␣
      ↪cost_centre_wbs_elment_order_no, and␣
      ↪cost_centre_wbs_element_order_no_description
```

```
# as this will not affect the outcome of our clustering analysis.
combined_pcard.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 183736 entries, 0 to 214775
Data columns (total 16 columns):
 #   Column                                   Non-Null Count   Dtype
---  ------                                   --------------   -----
 0   division                                 183736 non-null  object
 1   batch_transaction_id                     183736 non-null  object
 2   transaction_date                         183736 non-null  datetime64[ns]
 3   card_posting_dt                          183736 non-null  datetime64[ns]
 4   merchant_name                            183736 non-null  object
 5   transaction_amt                          183736 non-null  float64
 6   trx_currency                             183736 non-null  object
 7   original_amount                          183736 non-null  float64
 8   original_currency                        183736 non-null  object
 9   gl_account                               183736 non-null  object
 10  gl_account_description                   183736 non-null  object
 11  cost_centre_wbselement_orderno           183104 non-null  object
 12  cost_centre_wbselement_ordernodescription 183495 non-null  object
 13  merchant_type                            183716 non-null  float64
 14  merchant_type_description                183736 non-null  object
 15  purpose                                  182350 non-null  object
dtypes: datetime64[ns](2), float64(3), object(11)
memory usage: 23.8+ MB
```

```
[ ]: combined_pcard.nunique()
```

```
[ ]: division                                       58
     batch_transaction_id                       183720
     transaction_date                             1290
     card_posting_dt                               951
     merchant_name                               15217
     transaction_amt                             48568
     trx_currency                                    1
     original_amount                             47885
     original_currency                              22
     gl_account                                    303
     gl_account_description                        303
     cost_centre_wbselement_orderno               9621
     cost_centre_wbselement_ordernodescription    2612
     merchant_type                                 321
     merchant_type_description                     314
     purpose                                     94061
     dtype: int64
```

[ ]:

```python
#1.2.10
#Since we will be focusing mostly on the the Division and GL Account features␣
↪let us start by cleaning up the Division Column.

#  Strip leading and trailing spaces
combined_pcard['division'] = combined_pcard['division'].str.strip()

# Replace & with AND and then remove special characters (but first check if the␣
↪value is a string)
combined_pcard['division'] = combined_pcard['division'].apply(
    lambda x: re.sub(r'[^a-zA-Z0-9\s]', '', re.sub(r'&', 'AND', x))
)

# Convert to uppercase
combined_pcard['division'] = combined_pcard['division'].str.upper()

sorted(combined_pcard['division'].unique())
```

[ ]:

```
['311 TORONTO',
 'ACCOUNTING SERVICES',
 'AFFORDABLE HOUSING OFFICE',
 'CFO',
 'CHILDRENS SERVICES',
 'CITY CLERKS OFFICE',
 'CITY MANAGER',
 'CITY MANAGERS OFFICE',
 'CITY PLANNING',
 'CORPORATE CONTRACTS',
 'CORPORATE FINANCE',
 'CORPORATE SECURITY',
 'COURT SERVICES',
 'DEPUTY CITY MGR AND CFO',
 'DEPUTY CITY MGR INTERNAL SERVICES',
 'ECONOMIC DEVELOPMENT AND CULTURE',
 'EMERGENCY MEDICAL SERVICES',
 'EMPLOYMENT AND SOCIAL SERVICES',
 'ENGINEERING AND CONSTRUCTION SERVICES',
 'ENVIRONMENT AND ENERGY',
 'ENVIRONMENT AND ENERGY OFFICE',
 'EXECUTIVE MANAGEMENT',
 'FACILITIES MANAGEMENT',
 'FACILITIES MANAGEMENT DIVISON',
 'FINANCE AND ADMINISTRATION',
 'FINANCIAL PLANNING',
 'FIRE SERVICES',
 'FLEET SERVICES',
 'HUMAN RESOURCES',
 'INFORMATION AND TECHNOLOGY',
 'INTERNAL AUDIT',
 'LEGAL SERVICES',
 'LONG TERM CARE HOMES',
 'LONG TERM CARE HOMES AND SERVICES',
 'MUNICIPAL LICENSING AND STANDARDS',
 'OFFICE OF EMERGENCY MANAGEMENT',
 'PARKS FORESTRY AND RECREATION',
 'PENSION PAYROLL AND EMPLOYEE BENEFITS',
 'POLICY PLANNING FINANCE AND ADMINISTRATION',
 'PUBLIC HEALTH',
 'PURCHASING AND MATERIALS MANAGEMENT',
 'REAL ESTATE SERVICES',
 'REVENUE SERVICES',
 'SHELTER SUPPORT AND HOUSING ADMINISTRATION',
 'SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION',
 'SOLID WASTE MANAGEMENT',
 'STRATEGIC AND CORPORATE POLICY',
 'STRATEGIC COMMUNICATIONS',
 'TORONTO BUILDING',
 'TORONTO PARAMEDIC SERVICES',
 'TORONTO WATER',
 'TRANSPORTATION SERVICES',
 'TREASURER']
```

[ ]:

```
#1.2.11
#For divisions with minimal data or significance, we can group them together␣
 ↪for clarity.
#This will require us to get more information from the Toronto website on how␣
 ↪we can best group the divisions.

#1.2.11.1
#For this analysis we have used the latest Dec 2022 Divisions as highlighted in␣
 ↪the Business Expense report.

import requests
import pdfplumber


url = "https://www.toronto.ca/wp-content/uploads/2023/05/
 ↪9833-Business-Expense-Division-Dec-2022-vFinal-FINAL.pdf"
local_filename = "temp_division_pdf_file.pdf"

# Download the file
response = requests.get(url, stream=True)
with open(local_filename, 'wb') as f:
    for chunk in response.iter_content(chunk_size=8192):
```

```python
        f.write(chunk)

# Open the downloaded file with pdfplumber
with pdfplumber.open(local_filename) as pdf:
    # Extract tables from all pages
    all_tables = [page.extract_table() for page in pdf.pages if page.
 ↪extract_table() is not None]

# Convert the tables to DataFrames
dfs = [pd.DataFrame(table[1:], columns=table[0]) for table in all_tables]

# Remove the temporary PDF file
os.remove(local_filename)

unique_divisions = [x for x in dfs[0]['Division'].str.upper().unique() if x is␣
 ↪not None]
sorted_divisions = sorted(unique_divisions)
sorted_divisions
```

```
[ ]: ['ACCOUNTING SERVICES',
      'CAPITAL MARKET',
      "CHILDREN'S SERVICES",
      "CITY CLERK'S OFFICE",
      'CITY PLANNING',
      'CORPORATE REAL ESTATE MANAGEMENT',
      'COURT SERVICES',
      'CUSTOMER EXPERIENCE (311 TORONTO)',
      'ECONOMIC DEVELOPMENT AND CULTURE',
      'ENGINEERING & CONSTRUCTION SERVICES',
      'ENVIRONMENT & CLIMATE',
      'EXECUTIVE ADMINISTRATION',
      'FINANCIAL CONTROL AND PROCESS IMPROVEMENT',
      'FLEET SERVICES',
      'HOUSING SECRETARIAT',
      'INDIGENOUS AFFAIRS OFFICE',
      'INSURANCE & RISK MANAGEMENT',
      'INTERNAL AUDIT',
      'LEGAL SERVICES',
      'MUNICIPAL LICENSING AND STANDARDS',
      'OFFICE OF EMERGENCY MANAGEMENT',
      'OFFICE OF THE CHIEF OF STAFF',
      'OFFICE OF THE CISO',
      'OFFICE OF THE CONTROLLER',
      'OFFICE OF THE DEPUTY CITY MANAGER',
      'PARKS, FORESTRY & RECREATION',
      'PENSION, PAYROLL & EMPLOYEE BENEFITS',
      'PEOPLE & EQUITY',
      'POLICY, PLANNING, FINANCE & ADMINISTRATION',
      'PURCHASING & MATERIALS MANAGEMENT',
      'REVENUE SERVICES',
      'SENIORS SERVICES AND LONG-TERM CARE',
      'SHELTER, SUPPORT AND HOUSING ADMINISTRATION',
      'SOCIAL DEVELOPMENT, FINANCE AND ADMINISTRATION',
      'SOLID WASTE MANAGEMENT SERVICES',
      'STRATEGIC PARTNERSHIPS',
      'STRATEGIC PUBLIC & EMPLOYEE COMMUNICATIONS',
      'TECHNOLOGY SERVICES',
      'TORONTO BUILDING',
      'TORONTO EMPLOYMENT & SOCIAL SERVICES',
      'TORONTO FIRE SERVICES',
      'TORONTO PARAMEDIC SERVICES',
      'TORONTO PUBLIC HEALTH',
      'TORONTO WATER',
      'TOTAL',
      'TRANSIT EXPANSION OFFICE']
```

[ ]:

```
#1.2.11.2
#Based on the output of 1.2.11.1 let us do mapping of names to standardize the␣
 ↪division column.

name_mapping = {
    '311 TORONTO':'CUSTOMER EXPERIENCE (311 TORONTO)',
    'AFFORDABLE HOUSING OFFICE':'SHELTER SUPPORT AND HOUSING ADMINISTRATION',
    'CFO':'POLICY PLANNING FINANCE AND ADMINISTRATION',
    'DEPUTY CITY MGR AND CFO':'OFFICE OF THE DEPUTY CITY MANAGER',
    'DEPUTY CITY MGR INTERNAL SERVICES':'OFFICE OF THE DEPUTY CITY MANAGER',
    'CITY MANAGER':'CITY MANAGERS OFFICE',
    'CORPORATE CONTRACTS':'CORPORATE REAL ESTATE MANAGEMENT',
    'CORPORATE FINANCE':'OFFICE OF THE DEPUTY CITY MANAGER',
    'CORPORATE SECURITY':'CORPORATE REAL ESTATE MANAGEMENT',
    'EMERGENCY MEDICAL SERVICES':'OFFICE OF EMERGENCY MANAGEMENT',
    'EMPLOYMENT AND SOCIAL SERVICES': 'SOCIAL DEVELOPMENT FINANCE AND␣
 ↪ADMINISTRATION',
    'ENVIRONMENT AND ENERGY':'ENVIRONMENT AND CLIMATE',
    'ENVIRONMENT AND ENERGY OFFICE':'ENVIRONMENT AND CLIMATE',
    'FACILITIES MANAGEMENT': 'CORPORATE REAL ESTATE MANAGEMENT',
    'FACILITIES MANAGEMENT DIVISON':'CORPORATE REAL ESTATE MANAGEMENT',
    'EXECUTIVE MANAGEMENT':'CITY MANAGERS OFFICE',
    'FINANCE AND ADMINISTRATION':'POLICY PLANNING FINANCE AND ADMINISTRATION',
    'FINANCIAL PLANNING':'POLICY PLANNING FINANCE AND ADMINISTRATION',
    'FIRE SERVICES':'TORONTO FIRE SERVICES',
    'HUMAN RESOURCES': 'PEOPLE & EQUITY',
    'INFORMATION AND TECHNOLOGY':'TECHNOLOGY SERVICES',
    'LONG TERM CARE HOMES':'SENIORS SERVICES AND LONGTERM CARE',
    'LONG TERM CARE HOMES AND SERVICES': 'SENIORS SERVICES AND LONGTERM CARE',
    'PUBLIC HEALTH': 'TORONTO PUBLIC HEALTH',
    'REAL ESTATE SERVICES': 'CORPORATE REAL ESTATE MANAGEMENT',
    'SOLID WASTE MANAGEMENT':'SOLID WASTE MANAGEMENT SERVICES',
    'STRATEGIC AND CORPORATE POLICY':'CITY MANAGERS OFFICE',
    'STRATEGIC COMMUNICATIONS':'CITY MANAGERS OFFICE',
    'TREASURER':'POLICY PLANNING FINANCE AND ADMINISTRATION'
}

# Replace names using the mapping
combined_pcard['division'] = combined_pcard['division'].replace(name_mapping)

combined_pcard['division'].unique()
```

```
[ ]: array(['TORONTO PUBLIC HEALTH', 'ECONOMIC DEVELOPMENT AND CULTURE',
            'PARKS FORESTRY AND RECREATION', 'TORONTO FIRE SERVICES',
            'TRANSPORTATION SERVICES', 'TORONTO WATER',
            'SENIORS SERVICES AND LONGTERM CARE', 'TORONTO PARAMEDIC SERVICES',
            'SHELTER SUPPORT AND HOUSING ADMINISTRATION', 'CITY CLERKS OFFICE',
            'MUNICIPAL LICENSING AND STANDARDS',
            'SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION',
            'CORPORATE REAL ESTATE MANAGEMENT',
            'POLICY PLANNING FINANCE AND ADMINISTRATION', 'FLEET SERVICES',
            'SOLID WASTE MANAGEMENT SERVICES', 'TECHNOLOGY SERVICES',
            'CITY PLANNING', 'LEGAL SERVICES', 'CHILDRENS SERVICES',
            'ACCOUNTING SERVICES', 'ENGINEERING AND CONSTRUCTION SERVICES',
            'PURCHASING AND MATERIALS MANAGEMENT', 'CITY MANAGERS OFFICE',
            'TORONTO BUILDING', 'REVENUE SERVICES', 'ENVIRONMENT AND CLIMATE',
            'OFFICE OF THE DEPUTY CITY MANAGER',
            'PENSION PAYROLL AND EMPLOYEE BENEFITS',
            'OFFICE OF EMERGENCY MANAGEMENT', 'COURT SERVICES',
            'PEOPLE & EQUITY', 'CUSTOMER EXPERIENCE (311 TORONTO)',
            'INTERNAL AUDIT'], dtype=object)
```

```
[ ]: combined_pcard.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 183736 entries, 0 to 214775
Data columns (total 16 columns):
 #   Column                                    Non-Null Count   Dtype
---  ------                                    --------------   -----
 0   division                                  183736 non-null  object
 1   batch_transaction_id                      183736 non-null  object
 2   transaction_date                          183736 non-null  datetime64[ns]
 3   card_posting_dt                           183736 non-null  datetime64[ns]
 4   merchant_name                             183736 non-null  object
 5   transaction_amt                           183736 non-null  float64
 6   trx_currency                              183736 non-null  object
 7   original_amount                           183736 non-null  float64
 8   original_currency                         183736 non-null  object
 9   gl_account                                183736 non-null  object
 10  gl_account_description                    183736 non-null  object
 11  cost_centre_wbselement_orderno            183104 non-null  object
 12  cost_centre_wbselement_ordernodescription 183495 non-null  object
 13  merchant_type                             183716 non-null  float64
 14  merchant_type_description                 183736 non-null  object
 15  purpose                                   182350 non-null  object
dtypes: datetime64[ns](2), float64(3), object(11)
memory usage: 23.8+ MB
```

1.3.2 2. Division Profiling

- 2.1 Analyze expenditure amounts by division and Generate essential statistics for each Division's expenses such as measures of centrality, dispersion, and frequency.
- 2.2 Examine other potential characteristics including the diversity of the Cost Centres within each Division, usage of GL Accounts, choice of currencies, among others.
- 2.3 Aggregate the data entirely or present views by month or quarter.

Note: For divisions with minimal data or significance, group them together for clarity.

2.1 Characterize expenditure amounts by division and generate essential statistics for each Division's expenses such as measures of centrality, dispersion, and frequency. 2.1.1 Centrality, Dispersion, and Frequency Metrics for each Division's expenses:

[ ]:

```
# Grouping by division and aggregating transaction amounts
division_stats = combined_pcard.groupby('division')['transaction_amt'].
 ↪agg(['mean', 'median', 'std', 'count', 'min', 'max']).reset_index()
division_stats['range'] = division_stats['max'] - division_stats['min']

# Rounding to 2 decimal places
cols_to_round = ['mean', 'median', 'std', 'min', 'max', 'range']  # List of␣
 ↪columns to round
for col in cols_to_round:
    division_stats[col] = division_stats[col].round(2)

division_stats.sort_values(by='count')
```

[ ]:

```
                                   division    mean  median      std  \
12                           INTERNAL AUDIT  293.79  293.79      NaN
6                            COURT SERVICES  842.28  263.04  1102.37
7          CUSTOMER EXPERIENCE (311 TORONTO)  191.32  169.34   363.13
18       PENSION PAYROLL AND EMPLOYEE BENEFITS  447.04  450.87   420.37
21        PURCHASING AND MATERIALS MANAGEMENT  448.65  286.17   502.46
19                          PEOPLE & EQUITY  201.52   65.54   386.83
```

```
16            OFFICE OF THE DEPUTY CITY MANAGER   328.11  143.70   570.55
15               OFFICE OF EMERGENCY MANAGEMENT   372.25   92.94   835.95
28                             TORONTO BUILDING   613.27  247.69   857.18
0                           ACCOUNTING SERVICES   732.83  804.19   730.37
22                             REVENUE SERVICES   502.82  131.06  2593.37
27                          TECHNOLOGY SERVICES   312.73  136.70   601.50
20    POLICY PLANNING FINANCE AND ADMINISTRATION   500.14  210.87   757.78
1                            CHILDRENS SERVICES   651.46  367.82   852.27
4                                 CITY PLANNING   508.32  129.94   875.40
3                          CITY MANAGERS OFFICE   175.43   76.32   292.87
2                            CITY CLERKS OFFICE   456.69  187.45   769.35
10                      ENVIRONMENT AND CLIMATE   442.94  169.50   875.30
14            MUNICIPAL LICENSING AND STANDARDS   193.70   50.00   480.15
33                      TRANSPORTATION SERVICES   214.00   98.50   341.93
9          ENGINEERING AND CONSTRUCTION SERVICES    55.53   30.55   217.06
11                               FLEET SERVICES  2711.51  491.00  3693.09
30                    TORONTO PARAMEDIC SERVICES   295.32   62.28  1371.50
24    SHELTER SUPPORT AND HOUSING ADMINISTRATION   805.01  265.21  2090.86
26               SOLID WASTE MANAGEMENT SERVICES   205.59   90.39   355.12
29                        TORONTO FIRE SERVICES   293.61   77.20   613.57
31                        TORONTO PUBLIC HEALTH   387.27  145.00   733.21
23             SENIORS SERVICES AND LONGTERM CARE   396.27  156.50   654.44
25  SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION   235.06   50.84   620.60
13                               LEGAL SERVICES    93.71   17.00   251.78
32                                TORONTO WATER   203.36  104.37   334.34
5               CORPORATE REAL ESTATE MANAGEMENT   199.47   83.87   357.30
8               ECONOMIC DEVELOPMENT AND CULTURE   363.97   79.63  1164.00
17                PARKS FORESTRY AND RECREATION   199.38   81.81   341.98
```

```
    count      min       max     range
12      1   293.79    293.79      0.00
6      18    19.66   3384.35   3364.69
7      20  -734.50    827.14   1561.64
18     28   -89.27   1695.00   1784.27
21    110  -318.66   2687.12   3005.78
19    112  -399.80   1199.40   1599.20
16    114 -1400.00   3467.84   4867.84
15    151  -446.80   7289.63   7736.43
28    188 -2970.67   2998.00   5968.67
0     218 -3744.30   3744.30   7488.60
22    240  -777.02  39493.50  40270.52
27    241   -62.09   3164.00   3226.09
20    336 -1017.00   5210.00   6227.00
1     345 -3842.00   4294.00   8136.00
4     350  -371.52   6107.56   6479.08
3     455  -899.54   2541.79   3441.33
2     587 -1000.00   6942.70   7942.70
```

```
10    701    -669.53   6667.00   7336.53
14    904    -599.88   8859.20   9459.08
33   1397    -455.82   2841.09   3296.91
9    1426   -3334.26   3334.26   6668.52
11   2005   -4249.25  10000.00  14249.25
30   2075   -2500.00  41205.75  43705.75
24   3085  -16150.00  25000.00  41150.00
26   3190   -1050.90   4514.35   5565.25
29   4319   -2272.69  18958.50  21231.19
31   4936   -3000.00   9945.70  12945.70
23   5298   -5022.49   7079.45  12101.94
25   5458   -4746.00   6780.00  11526.00
13   5821    -785.35   6068.52   6853.87
32  11709   -2807.00   3000.00   5807.00
5   14715   -2527.71  15010.37  17538.08
8   14846   -4234.54  61842.37  66076.91
17  98337   -3000.00   6162.12   9162.12
```

2.2 Examine other potential characteristics including the diversity of the Cost Centres within each Division, usage of GL Accounts, choice of currencies, among others.

2.2.1 Diversity/Concentration Metrics:

[ ]:

```
# Calculate total unique values for each category across the entire dataset
total_unique_cost_centres = combined_pcard['cost_centre_wbselement_orderno'].
 ↪nunique()
total_unique_gl_accounts = combined_pcard['gl_account'].nunique()
total_unique_original_currency = combined_pcard['original_currency'].nunique()
total_unique_merchant_type = combined_pcard['merchant_type'].nunique()
total_unique_purpose = combined_pcard['purpose'].nunique()

# Diversity/Concentration Metrics by Division
def get_diversity_df(groupby_column, total_unique):
    counts = combined_pcard.groupby('division')[groupby_column].nunique()
    percentages = ((counts / total_unique) * 100).round(1)
    df = pd.DataFrame({
        'Count': counts,
        'Percentage of Total (%)': percentages
    }).sort_values(by='Count', ascending=False)
    return df

div_cost_centre_df = get_diversity_df('cost_centre_wbselement_orderno',␣
 ↪total_unique_cost_centres)
div_gl_account_df = get_diversity_df('gl_account', total_unique_gl_accounts)
div_unique_original_currency = get_diversity_df('original_currency',␣
 ↪total_unique_original_currency)
```

```
div_unique_merchant_type = get_diversity_df('merchant_type',␣
 ↪total_unique_merchant_type)
div_purpose_df = get_diversity_df('purpose', total_unique_purpose)


print("Diversity/Concentration Metrics for each Division:")
print("\nNumber of unique Cost Centres:")
display(div_cost_centre_df)
print(" ")
print("\nNumber of unique GL Accounts:")
display(div_gl_account_df)
print(" ")
print("\nNumber of unique Currencies:")
display(div_unique_original_currency)
print(" ")
print("\nNumber of unique Merchant Types:")
display(div_unique_merchant_type)
print(" ")
print("\nNumber of unique Purposes:")
display(div_purpose_df)
```

```
Diversity/Concentration Metrics for each Division:

Number of unique Cost Centres:

                                               Count  Percentage of Total (%)
division
CORPORATE REAL ESTATE MANAGEMENT                7010                     72.9
PARKS FORESTRY AND RECREATION                   1532                     15.9
ECONOMIC DEVELOPMENT AND CULTURE                 172                      1.8
TORONTO PUBLIC HEALTH                            125                      1.3
TORONTO WATER                                    105                      1.1
SENIORS SERVICES AND LONGTERM CARE                98                      1.0
SOLID WASTE MANAGEMENT SERVICES                   84                      0.9
LEGAL SERVICES                                    76                      0.8
SHELTER SUPPORT AND HOUSING ADMINISTRATION        74                      0.8
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION     65                      0.7
TORONTO PARAMEDIC SERVICES                        59                      0.6
TRANSPORTATION SERVICES                           35                      0.4
POLICY PLANNING FINANCE AND ADMINISTRATION        32                      0.3
MUNICIPAL LICENSING AND STANDARDS                 27                      0.3
CITY CLERKS OFFICE                                25                      0.3
TORONTO FIRE SERVICES                             25                      0.3
OFFICE OF THE DEPUTY CITY MANAGER                 20                      0.2
OFFICE OF EMERGENCY MANAGEMENT                    19                      0.2
ENVIRONMENT AND CLIMATE                           16                      0.2
ACCOUNTING SERVICES                               16                      0.2
PURCHASING AND MATERIALS MANAGEMENT               11                      0.1
```

```
REVENUE SERVICES                                  10                      0.1
CITY PLANNING                                      9                      0.1
CITY MANAGERS OFFICE                               9                      0.1
ENGINEERING AND CONSTRUCTION SERVICES              8                      0.1
FLEET SERVICES                                     7                      0.1
COURT SERVICES                                     6                      0.1
PENSION PAYROLL AND EMPLOYEE BENEFITS              6                      0.1
TECHNOLOGY SERVICES                                5                      0.1
TORONTO BUILDING                                   5                      0.1
CHILDRENS SERVICES                                 3                      0.0
PEOPLE & EQUITY                                    2                      0.0
CUSTOMER EXPERIENCE (311 TORONTO)                  2                      0.0
INTERNAL AUDIT                                     1                      0.0


Number of unique GL Accounts:

                                               Count  Percentage of Total (%)
division
PARKS FORESTRY AND RECREATION                    219                     72.3
ECONOMIC DEVELOPMENT AND CULTURE                 147                     48.5
TORONTO WATER                                    138                     45.5
SHELTER SUPPORT AND HOUSING ADMINISTRATION       131                     43.2
SOLID WASTE MANAGEMENT SERVICES                  125                     41.3
TORONTO PUBLIC HEALTH                            106                     35.0
TRANSPORTATION SERVICES                          101                     33.3
CORPORATE REAL ESTATE MANAGEMENT                  99                     32.7
TORONTO FIRE SERVICES                             95                     31.4
SENIORS SERVICES AND LONGTERM CARE                83                     27.4
TORONTO PARAMEDIC SERVICES                        77                     25.4
MUNICIPAL LICENSING AND STANDARDS                 77                     25.4
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION     64                     21.1
CITY CLERKS OFFICE                                49                     16.2
ENVIRONMENT AND CLIMATE                           45                     14.9
OFFICE OF EMERGENCY MANAGEMENT                    41                     13.5
REVENUE SERVICES                                  40                     13.2
FLEET SERVICES                                    37                     12.2
OFFICE OF THE DEPUTY CITY MANAGER                 28                      9.2
TORONTO BUILDING                                  28                      9.2
POLICY PLANNING FINANCE AND ADMINISTRATION        24                      7.9
PURCHASING AND MATERIALS MANAGEMENT               23                      7.6
ACCOUNTING SERVICES                               22                      7.3
CITY MANAGERS OFFICE                              21                      6.9
TECHNOLOGY SERVICES                               18                      5.9
LEGAL SERVICES                                    18                      5.9
CITY PLANNING                                     17                      5.6
CHILDRENS SERVICES                                13                      4.3
COURT SERVICES                                    10                      3.3
```

```
ENGINEERING AND CONSTRUCTION SERVICES              9                        3.0
PENSION PAYROLL AND EMPLOYEE BENEFITS              7                        2.3
CUSTOMER EXPERIENCE (311 TORONTO)                  6                        2.0
PEOPLE & EQUITY                                    4                        1.3
INTERNAL AUDIT                                     1                        0.3
```

```
Number of unique Currencies:
                                               Count  Percentage of Total (%)
division
ECONOMIC DEVELOPMENT AND CULTURE                  20                     90.9
TORONTO PUBLIC HEALTH                              7                     31.8
PARKS FORESTRY AND RECREATION                      6                     27.3
CITY PLANNING                                      4                     18.2
TORONTO WATER                                      4                     18.2
FLEET SERVICES                                     3                     13.6
SENIORS SERVICES AND LONGTERM CARE                 3                     13.6
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION      3                     13.6
SOLID WASTE MANAGEMENT SERVICES                    3                     13.6
TORONTO FIRE SERVICES                              3                     13.6
TORONTO PARAMEDIC SERVICES                         3                     13.6
CITY MANAGERS OFFICE                               3                     13.6
CITY CLERKS OFFICE                                 3                     13.6
TORONTO BUILDING                                   2                      9.1
POLICY PLANNING FINANCE AND ADMINISTRATION         2                      9.1
TECHNOLOGY SERVICES                                2                      9.1
SHELTER SUPPORT AND HOUSING ADMINISTRATION         2                      9.1
REVENUE SERVICES                                   2                      9.1
PURCHASING AND MATERIALS MANAGEMENT                2                      9.1
ACCOUNTING SERVICES                                2                      9.1
PEOPLE & EQUITY                                    2                      9.1
CHILDRENS SERVICES                                 2                      9.1
OFFICE OF THE DEPUTY CITY MANAGER                  2                      9.1
OFFICE OF EMERGENCY MANAGEMENT                     2                      9.1
MUNICIPAL LICENSING AND STANDARDS                  2                      9.1
LEGAL SERVICES                                     2                      9.1
ENVIRONMENT AND CLIMATE                            2                      9.1
ENGINEERING AND CONSTRUCTION SERVICES              2                      9.1
CORPORATE REAL ESTATE MANAGEMENT                   2                      9.1
TRANSPORTATION SERVICES                            2                      9.1
PENSION PAYROLL AND EMPLOYEE BENEFITS              1                      4.5
INTERNAL AUDIT                                     1                      4.5
CUSTOMER EXPERIENCE (311 TORONTO)                  1                      4.5
COURT SERVICES                                     1                      4.5
```

```
Number of unique Merchant Types:

                                               Count  Percentage of Total (%)
division
ECONOMIC DEVELOPMENT AND CULTURE                 236                     73.5
PARKS FORESTRY AND RECREATION                    224                     69.8
TORONTO FIRE SERVICES                            160                     49.8
TORONTO WATER                                    159                     49.5
SHELTER SUPPORT AND HOUSING ADMINISTRATION       159                     49.5
TORONTO PUBLIC HEALTH                            152                     47.4
SENIORS SERVICES AND LONGTERM CARE               143                     44.5
CORPORATE REAL ESTATE MANAGEMENT                 142                     44.2
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION    117                     36.4
SOLID WASTE MANAGEMENT SERVICES                  111                     34.6
TORONTO PARAMEDIC SERVICES                       102                     31.8
CITY CLERKS OFFICE                                92                     28.7
TRANSPORTATION SERVICES                           90                     28.0
ENVIRONMENT AND CLIMATE                           84                     26.2
MUNICIPAL LICENSING AND STANDARDS                 79                     24.6
OFFICE OF EMERGENCY MANAGEMENT                    49                     15.3
REVENUE SERVICES                                  47                     14.6
POLICY PLANNING FINANCE AND ADMINISTRATION        43                     13.4
OFFICE OF THE DEPUTY CITY MANAGER                 42                     13.1
FLEET SERVICES                                    40                     12.5
LEGAL SERVICES                                    39                     12.1
TORONTO BUILDING                                  37                     11.5
CITY MANAGERS OFFICE                              36                     11.2
TECHNOLOGY SERVICES                               35                     10.9
CHILDRENS SERVICES                                35                     10.9
PURCHASING AND MATERIALS MANAGEMENT               35                     10.9
ACCOUNTING SERVICES                               33                     10.3
CITY PLANNING                                     29                      9.0
ENGINEERING AND CONSTRUCTION SERVICES             18                      5.6
COURT SERVICES                                    11                      3.4
CUSTOMER EXPERIENCE (311 TORONTO)                 10                      3.1
PEOPLE & EQUITY                                    9                      2.8
PENSION PAYROLL AND EMPLOYEE BENEFITS              7                      2.2
INTERNAL AUDIT                                     1                      0.3
```

```
Number of unique Purposes:

                                               Count  Percentage of Total (%)
division
PARKS FORESTRY AND RECREATION                  55494                     59.0
ECONOMIC DEVELOPMENT AND CULTURE                8704                      9.3
TORONTO WATER                                   7046                      7.5
CORPORATE REAL ESTATE MANAGEMENT                6702                      7.1
```

```
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION   4206                      4.5
SENIORS SERVICES AND LONGTERM CARE              3677                      3.9
SHELTER SUPPORT AND HOUSING ADMINISTRATION      2546                      2.7
SOLID WASTE MANAGEMENT SERVICES                 2321                      2.5
TORONTO FIRE SERVICES                           2280                      2.4
TORONTO PARAMEDIC SERVICES                      1032                      1.1
TORONTO PUBLIC HEALTH                            975                      1.0
TRANSPORTATION SERVICES                          910                      1.0
ENGINEERING AND CONSTRUCTION SERVICES            604                      0.6
CITY CLERKS OFFICE                               475                      0.5
ENVIRONMENT AND CLIMATE                          429                      0.5
MUNICIPAL LICENSING AND STANDARDS                403                      0.4
FLEET SERVICES                                   388                      0.4
CITY PLANNING                                    278                      0.3
CITY MANAGERS OFFICE                             250                      0.3
REVENUE SERVICES                                 213                      0.2
LEGAL SERVICES                                   210                      0.2
CHILDRENS SERVICES                               180                      0.2
TECHNOLOGY SERVICES                              172                      0.2
TORONTO BUILDING                                 155                      0.2
POLICY PLANNING FINANCE AND ADMINISTRATION       144                      0.2
ACCOUNTING SERVICES                              126                      0.1
OFFICE OF EMERGENCY MANAGEMENT                   111                      0.1
OFFICE OF THE DEPUTY CITY MANAGER                106                      0.1
PURCHASING AND MATERIALS MANAGEMENT              105                      0.1
CUSTOMER EXPERIENCE (311 TORONTO)                 20                      0.0
PENSION PAYROLL AND EMPLOYEE BENEFITS             20                      0.0
COURT SERVICES                                    17                      0.0
PEOPLE & EQUITY                                   13                      0.0
INTERNAL AUDIT                                     1                      0.0
```

2.3 Aggregate across the entire dataset or a view aggregated by month or quarter.

```
2.3.1 Average Monthly/Quarterly Expenditure Amount:
```

[ ]:

```python
# Extracting month and quarter from the transaction_date
combined_pcard['transaction_month'] = combined_pcard['transaction_date'].dt.month
combined_pcard['transaction_quarter'] = combined_pcard['transaction_date'].dt.quarter
combined_pcard['transaction_year'] = combined_pcard['transaction_date'].dt.year

# Monthly aggregation
monthly_expenses = combined_pcard.groupby(['division', 'transaction_month'])['transaction_amt'].mean().round(2).reset_index()

# Quarterly aggregation
```

```python
quarterly_expenses = combined_pcard.groupby(['division',␣
 ↪'transaction_quarter'])['transaction_amt'].mean().round(2).reset_index()

# Yearly aggregation
yearly_expenses = combined_pcard.groupby(['division',␣
 ↪'transaction_year'])['transaction_amt'].mean().round(2).reset_index()

# Pivot tables with NaN values replaced by empty strings
monthly_pivot = monthly_expenses.pivot_table(index='division',␣
 ↪columns='transaction_month', values='transaction_amt', aggfunc='mean').
 ↪fillna("")
print("Monthly Average by Division:")
display(monthly_pivot)
print("")
```

```
Monthly Average by Division:

transaction_month                                   1        2        3   \
division
ACCOUNTING SERVICES                              534.7   551.99   465.93
CHILDRENS SERVICES                               601.7   778.18   668.27
CITY CLERKS OFFICE                              377.63   405.27   346.46
CITY MANAGERS OFFICE                            139.88   178.22   322.46
CITY PLANNING                                   449.77   417.41   584.96
CORPORATE REAL ESTATE MANAGEMENT                197.26   181.98   199.86
COURT SERVICES                                 1862.57
CUSTOMER EXPERIENCE (311 TORONTO)                        142.13
ECONOMIC DEVELOPMENT AND CULTURE                384.31   342.07   347.66
ENGINEERING AND CONSTRUCTION SERVICES             50.2    44.46    54.64
ENVIRONMENT AND CLIMATE                         512.34   340.69   442.57
FLEET SERVICES                                 2268.69  3026.92  2030.62
INTERNAL AUDIT
LEGAL SERVICES                                   87.55    74.83   107.28
MUNICIPAL LICENSING AND STANDARDS               311.07    290.0   100.84
OFFICE OF EMERGENCY MANAGEMENT                  188.36   142.35  1115.32
OFFICE OF THE DEPUTY CITY MANAGER               315.74   288.74    78.01
PARKS FORESTRY AND RECREATION                   175.55   172.03   191.21
PENSION PAYROLL AND EMPLOYEE BENEFITS                    931.97   940.72
PEOPLE & EQUITY                                 1199.4    78.59   144.56
POLICY PLANNING FINANCE AND ADMINISTRATION      328.44   245.09   667.12
PURCHASING AND MATERIALS MANAGEMENT             337.66   451.06   337.91
REVENUE SERVICES                                248.42   323.46   300.99
SENIORS SERVICES AND LONGTERM CARE              483.39   453.67   376.53
SHELTER SUPPORT AND HOUSING ADMINISTRATION      964.63  1145.79   894.02
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION   225.54   276.91   270.24
SOLID WASTE MANAGEMENT SERVICES                 221.51   217.79    219.4
TECHNOLOGY SERVICES                             475.79   227.33   146.63
TORONTO BUILDING                                169.97   611.58   807.34
TORONTO FIRE SERVICES                           269.64   251.24   281.04
TORONTO PARAMEDIC SERVICES                      930.86   440.64   225.52
TORONTO PUBLIC HEALTH                           439.11    337.5   355.65
TORONTO WATER                                   206.37   205.59   206.55
TRANSPORTATION SERVICES                         153.77   178.55   200.14

transaction_month                                   4       5        6   \
division
ACCOUNTING SERVICES                              956.2  838.83   693.98
CHILDRENS SERVICES                               345.7  608.53   566.68
CITY CLERKS OFFICE                              358.99  485.94   660.16
CITY MANAGERS OFFICE                            171.53  114.26    66.87
CITY PLANNING                                   356.59  389.05   601.41
CORPORATE REAL ESTATE MANAGEMENT                205.33  211.39   204.45
COURT SERVICES                                  1107.4  502.05    93.77
CUSTOMER EXPERIENCE (311 TORONTO)               130.76            36.58
ECONOMIC DEVELOPMENT AND CULTURE                381.61  246.72   366.09
ENGINEERING AND CONSTRUCTION SERVICES            71.31   58.79    55.35
ENVIRONMENT AND CLIMATE                         394.65  359.49   321.56
FLEET SERVICES                                 2435.57  726.57   592.31
INTERNAL AUDIT
LEGAL SERVICES                                   76.12    92.7   105.98
MUNICIPAL LICENSING AND STANDARDS               228.59  244.27   252.58
OFFICE OF EMERGENCY MANAGEMENT                  202.22  154.72   990.85
OFFICE OF THE DEPUTY CITY MANAGER               232.89  286.04    483.5
PARKS FORESTRY AND RECREATION                   188.54  192.54   212.11
PENSION PAYROLL AND EMPLOYEE BENEFITS           488.63  492.12   388.68
PEOPLE & EQUITY                                 101.04  988.75  1124.35
POLICY PLANNING FINANCE AND ADMINISTRATION      593.47  979.64   291.44
PURCHASING AND MATERIALS MANAGEMENT             304.24  771.64   146.73
REVENUE SERVICES                               1897.68  482.08   330.07
SENIORS SERVICES AND LONGTERM CARE              343.94  347.95   351.07
SHELTER SUPPORT AND HOUSING ADMINISTRATION     1011.44  732.44   706.13
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION   254.16  195.16   205.27
SOLID WASTE MANAGEMENT SERVICES                 219.57  236.42   225.87
TECHNOLOGY SERVICES                             349.37  272.17   577.36
TORONTO BUILDING                                 326.2   573.5   549.38
TORONTO FIRE SERVICES                           270.98  269.49    307.9
TORONTO PARAMEDIC SERVICES                      290.33  214.44   179.34
TORONTO PUBLIC HEALTH                           382.82  426.95   344.85
TORONTO WATER                                   198.97  201.56    202.7
TRANSPORTATION SERVICES                         262.53  250.62   247.02

transaction_month                                  7        8        9   \
division
ACCOUNTING SERVICES                            821.57   320.39   554.25
CHILDRENS SERVICES                             761.47   980.22   655.01
CITY CLERKS OFFICE                             761.88   268.94   438.38
CITY MANAGERS OFFICE                           311.76   214.11   107.18
CITY PLANNING                                  384.99   598.61   166.45
CORPORATE REAL ESTATE MANAGEMENT               187.19   195.75   210.15
COURT SERVICES
CUSTOMER EXPERIENCE (311 TORONTO)              465.53            195.28
ECONOMIC DEVELOPMENT AND CULTURE               361.73   319.74   472.17
ENGINEERING AND CONSTRUCTION SERVICES           52.77     40.3    61.78
ENVIRONMENT AND CLIMATE                        458.46   477.35   575.23
FLEET SERVICES                                 457.54   5642.1  5211.09
INTERNAL AUDIT                                                   293.79
LEGAL SERVICES                                   90.6    94.39     95.9
MUNICIPAL LICENSING AND STANDARDS               59.23   135.12   338.35
OFFICE OF EMERGENCY MANAGEMENT                 398.58   450.87    364.0
OFFICE OF THE DEPUTY CITY MANAGER              819.25   1000.0    191.5
PARKS FORESTRY AND RECREATION                  215.66   210.85    235.1
PENSION PAYROLL AND EMPLOYEE BENEFITS          149.27   480.63   -89.27
PEOPLE & EQUITY                                        -374.81
POLICY PLANNING FINANCE AND ADMINISTRATION     408.25   450.09   127.16
PURCHASING AND MATERIALS MANAGEMENT            865.12   451.46   604.62
REVENUE SERVICES                                303.6   343.99   328.55
SENIORS SERVICES AND LONGTERM CARE             423.22   356.97   341.21
SHELTER SUPPORT AND HOUSING ADMINISTRATION      458.7   456.57   574.09
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION  298.41   384.01   142.04
SOLID WASTE MANAGEMENT SERVICES                201.02   180.21   134.42
TECHNOLOGY SERVICES                            105.71   106.91   151.06
TORONTO BUILDING                               513.05  1047.37   435.08
TORONTO FIRE SERVICES                          288.56   292.33   238.63
TORONTO PARAMEDIC SERVICES                     206.72    234.1   254.55
TORONTO PUBLIC HEALTH                          341.57   497.49   394.87
TORONTO WATER                                  207.91   193.39   214.83
TRANSPORTATION SERVICES                        208.93   151.27   270.63

transaction_month                                     10       11       12
division
ACCOUNTING SERVICES                             799.84   363.92   537.31
CHILDRENS SERVICES                              1396.4    591.9   511.68
CITY CLERKS OFFICE                              401.86   490.64   503.77
CITY MANAGERS OFFICE                            580.75   185.45   305.38
CITY PLANNING                                   654.42   840.14    948.7
CORPORATE REAL ESTATE MANAGEMENT                206.39   214.36   179.83
COURT SERVICES                                           894.88    19.66
CUSTOMER EXPERIENCE (311 TORONTO)               275.08     1.71   272.16
ECONOMIC DEVELOPMENT AND CULTURE                385.91   341.19    454.0
ENGINEERING AND CONSTRUCTION SERVICES            41.45    91.62   105.87
ENVIRONMENT AND CLIMATE                         822.55    306.0   578.14
FLEET SERVICES                                 4230.27  1894.93  1879.97
INTERNAL AUDIT
LEGAL SERVICES                                   84.72   115.06   126.82
```

```
MUNICIPAL LICENSING AND STANDARDS               227.81   178.69   135.87
OFFICE OF EMERGENCY MANAGEMENT                 1234.04   684.64   932.85
OFFICE OF THE DEPUTY CITY MANAGER               859.12   111.82   336.32
PARKS FORESTRY AND RECREATION                   188.94   188.43   232.12
PENSION PAYROLL AND EMPLOYEE BENEFITS
PEOPLE & EQUITY                                  78.05            1199.4
POLICY PLANNING FINANCE AND ADMINISTRATION      296.76   353.51   140.86
PURCHASING AND MATERIALS MANAGEMENT             515.21   347.27   367.02
REVENUE SERVICES                                 432.0   354.85   274.83
SENIORS SERVICES AND LONGTERM CARE               391.2   475.11   414.44
SHELTER SUPPORT AND HOUSING ADMINISTRATION      559.52   713.12  1081.76
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION   241.16   244.13   200.24
SOLID WASTE MANAGEMENT SERVICES                 194.14    177.8   176.02
TECHNOLOGY SERVICES                             200.61   180.65   415.26
TORONTO BUILDING                                849.65   896.75   811.93
TORONTO FIRE SERVICES                            253.5   404.16   438.78
TORONTO PARAMEDIC SERVICES                      301.18   134.98   178.94
TORONTO PUBLIC HEALTH                           348.17   360.59   469.13
TORONTO WATER                                   197.55   200.95   203.23
TRANSPORTATION SERVICES                         198.41   200.45    210.9
```

```
[ ]: quarterly_pivot = quarterly_expenses.pivot_table(index='division',␣
       ↪columns='transaction_quarter', values='transaction_amt', aggfunc='mean').
       ↪fillna("")
     print("\nQuarterly Average by Division:")
     display(quarterly_pivot)
     print("")
```

```
Quarterly Average by Division:

transaction_quarter                                  1        2        3  \
division
ACCOUNTING SERVICES                             520.19    836.7   603.19
CHILDRENS SERVICES                              685.86   498.79   743.26
CITY CLERKS OFFICE                              376.32   511.16   481.95
CITY MANAGERS OFFICE                             225.1   108.72   203.04
CITY PLANNING                                   492.23   461.08   329.75
CORPORATE REAL ESTATE MANAGEMENT                 192.8   207.02    197.6
COURT SERVICES                                 1862.57   500.64
CUSTOMER EXPERIENCE (311 TORONTO)               142.13    99.37   285.37
ECONOMIC DEVELOPMENT AND CULTURE                357.06   329.74    393.4
ENGINEERING AND CONSTRUCTION SERVICES            49.67    62.12    51.24
ENVIRONMENT AND CLIMATE                         434.72   356.91   492.11
FLEET SERVICES                                 2446.58  1318.36  4517.54
INTERNAL AUDIT                                                    293.79
LEGAL SERVICES                                   91.15    89.44    93.77
```

```
MUNICIPAL LICENSING AND STANDARDS              212.95   240.2   122.89
OFFICE OF EMERGENCY MANAGEMENT                 223.77  268.73   384.24
OFFICE OF THE DEPUTY CITY MANAGER              197.02  345.34    604.3
PARKS FORESTRY AND RECREATION                  179.99  198.48   218.19
PENSION PAYROLL AND EMPLOYEE BENEFITS          934.89  467.97   207.39
PEOPLE & EQUITY                                115.77  578.79  -374.81
POLICY PLANNING FINANCE AND ADMINISTRATION     411.87  714.78   325.45
PURCHASING AND MATERIALS MANAGEMENT            377.99   429.8   566.35
REVENUE SERVICES                               295.76  790.19   326.42
SENIORS SERVICES AND LONGTERM CARE             433.82  347.85   372.53
SHELTER SUPPORT AND HOUSING ADMINISTRATION    1003.47  808.83   506.51
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION  259.29  216.34   242.86
SOLID WASTE MANAGEMENT SERVICES                219.55  227.34   168.78
TECHNOLOGY SERVICES                            306.46  417.22   125.18
TORONTO BUILDING                               541.41  492.47   640.39
TORONTO FIRE SERVICES                          268.31  281.89   271.69
TORONTO PARAMEDIC SERVICES                     485.18  225.67   235.84
TORONTO PUBLIC HEALTH                          375.75  385.41   407.51
TORONTO WATER                                  206.18  201.09   205.18
TRANSPORTATION SERVICES                         177.0  253.37   217.84

transaction_quarter                                 4
division
ACCOUNTING SERVICES                            617.94
CHILDRENS SERVICES                            1126.92
CITY CLERKS OFFICE                             459.62
CITY MANAGERS OFFICE                           295.07
CITY PLANNING                                  832.13
CORPORATE REAL ESTATE MANAGEMENT               201.04
COURT SERVICES                                 676.08
CUSTOMER EXPERIENCE (311 TORONTO)              196.56
ECONOMIC DEVELOPMENT AND CULTURE               390.33
ENGINEERING AND CONSTRUCTION SERVICES           63.12
ENVIRONMENT AND CLIMATE                         526.2
FLEET SERVICES                                2962.45
INTERNAL AUDIT
LEGAL SERVICES                                 104.63
MUNICIPAL LICENSING AND STANDARDS              180.87
OFFICE OF EMERGENCY MANAGEMENT                1033.47
OFFICE OF THE DEPUTY CITY MANAGER              485.64
PARKS FORESTRY AND RECREATION                  199.83
PENSION PAYROLL AND EMPLOYEE BENEFITS
PEOPLE & EQUITY                               1087.27
POLICY PLANNING FINANCE AND ADMINISTRATION     287.76
PURCHASING AND MATERIALS MANAGEMENT            389.66
REVENUE SERVICES                               356.31
SENIORS SERVICES AND LONGTERM CARE             428.08
SHELTER SUPPORT AND HOUSING ADMINISTRATION     786.02
```

```
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION   232.22
SOLID WASTE MANAGEMENT SERVICES                 182.42
TECHNOLOGY SERVICES                             255.31
TORONTO BUILDING                                851.29
TORONTO FIRE SERVICES                           359.88
TORONTO PARAMEDIC SERVICES                      209.83
TORONTO PUBLIC HEALTH                           388.85
TORONTO WATER                                   200.55
TRANSPORTATION SERVICES                          202.3
```

[ ]:

```
yearly_pivot = yearly_expenses.pivot_table(index='division',␣
 ↪columns='transaction_year', values='transaction_amt', aggfunc='mean').
 ↪fillna("")
print("\nYearly Average by Division:")
display(yearly_pivot)
```

```
Yearly Average by Division:

transaction_year                                 2014     2015     2016  \
division
ACCOUNTING SERVICES                                     721.84   608.14
CHILDRENS SERVICES                                      645.63   903.74
CITY CLERKS OFFICE                                      407.77   439.22
CITY MANAGERS OFFICE                                    227.75   226.44
CITY PLANNING                                           645.06   376.63
CORPORATE REAL ESTATE MANAGEMENT                92.01   183.63   200.87
COURT SERVICES                                          473.39
CUSTOMER EXPERIENCE (311 TORONTO)
ECONOMIC DEVELOPMENT AND CULTURE                88.91   345.61   334.61
ENGINEERING AND CONSTRUCTION SERVICES                    39.08    74.49
ENVIRONMENT AND CLIMATE                                 418.93   416.12
FLEET SERVICES                                         2699.61  2756.53
INTERNAL AUDIT                                                   293.79
LEGAL SERVICES                                           40.54   139.44
MUNICIPAL LICENSING AND STANDARDS                       122.88   138.93
OFFICE OF EMERGENCY MANAGEMENT                 186.53   163.89  1246.64
OFFICE OF THE DEPUTY CITY MANAGER                       372.99   218.84
PARKS FORESTRY AND RECREATION                  173.07   190.88   207.44
PENSION PAYROLL AND EMPLOYEE BENEFITS                   601.91   540.01
PEOPLE & EQUITY                                         1199.4    93.02
POLICY PLANNING FINANCE AND ADMINISTRATION              245.67   311.58
PURCHASING AND MATERIALS MANAGEMENT                     401.79   586.97
REVENUE SERVICES                              -112.86   384.07   338.28
SENIORS SERVICES AND LONGTERM CARE             656.21    450.5   428.88
SHELTER SUPPORT AND HOUSING ADMINISTRATION     499.44   536.08   791.25
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION           223.44   217.65
```

```
SOLID WASTE MANAGEMENT SERVICES                 70.81   183.37   172.89
TECHNOLOGY SERVICES                                     317.73   202.39
TORONTO BUILDING                                        619.11   480.65
TORONTO FIRE SERVICES                          386.64    274.9   287.23
TORONTO PARAMEDIC SERVICES                              187.66   208.32
TORONTO PUBLIC HEALTH                          150.34   370.35   363.57
TORONTO WATER                                  203.68   199.05   198.68
TRANSPORTATION SERVICES                                 237.22   221.95

transaction_year                                  2017     2018
division
ACCOUNTING SERVICES                             798.98  1035.89
CHILDRENS SERVICES                              811.99   332.71
CITY CLERKS OFFICE                              421.25   606.65
CITY MANAGERS OFFICE                            259.83    94.86
CITY PLANNING                                   447.96   805.94
CORPORATE REAL ESTATE MANAGEMENT                195.49   224.92
COURT SERVICES                                 2290.34    383.7
CUSTOMER EXPERIENCE (311 TORONTO)               237.55   105.48
ECONOMIC DEVELOPMENT AND CULTURE                418.13    331.4
ENGINEERING AND CONSTRUCTION SERVICES            69.07
ENVIRONMENT AND CLIMATE                          507.7   410.85
FLEET SERVICES                                 2796.81  2414.68
INTERNAL AUDIT
LEGAL SERVICES                                   237.8   270.57
MUNICIPAL LICENSING AND STANDARDS               226.91   328.81
OFFICE OF EMERGENCY MANAGEMENT                  470.65   419.51
OFFICE OF THE DEPUTY CITY MANAGER               596.74    213.5
PARKS FORESTRY AND RECREATION                   203.06   192.45
PENSION PAYROLL AND EMPLOYEE BENEFITS            234.3   759.92
PEOPLE & EQUITY                                 423.21
POLICY PLANNING FINANCE AND ADMINISTRATION      577.68   787.65
PURCHASING AND MATERIALS MANAGEMENT             370.23   492.57
REVENUE SERVICES                                323.42   820.49
SENIORS SERVICES AND LONGTERM CARE              317.07   365.83
SHELTER SUPPORT AND HOUSING ADMINISTRATION      859.76  1030.41
SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION   264.84   221.07
SOLID WASTE MANAGEMENT SERVICES                 217.46   244.72
TECHNOLOGY SERVICES                             370.84   312.94
TORONTO BUILDING                                786.83   644.51
TORONTO FIRE SERVICES                           301.71   322.66
TORONTO PARAMEDIC SERVICES                      360.05   559.08
TORONTO PUBLIC HEALTH                           413.71   387.69
TORONTO WATER                                   208.31   209.15
TRANSPORTATION SERVICES                         199.54   201.65
```

The dataset output above shows the yearly average employee expenditures for various divisions within the Toronto City government. Here's an interpretation of sample divisions:

- **Accounting Services:**
  - Expenditures have shown an increasing trend from 2014 to 2017, peaking in 2017 ($1035.89).
- **Children's Services:**
  - Expenditures vary with the highest in 2015 ($903.74) and a significant drop in 2018 ($332.71).
- **Economic Development and Culture:**
  - Expenditures vary, peaking in 2016 ($418.13) and showing fluctuations in other years.
- **Engineering and Construction Services:**
  - Expenditures vary, with noticeable amounts in 2015 ($74.49) and 2016 ($69.07).
- **Environment and Climate:**
  - Expenditures have fluctuations, with the highest in 2017 ($507.70) and variations in other years.
- **Fleet Services:**
  - Expenditures show fluctuations, with relatively stable amounts in 2015 to 2017 and a slight decrease in 2018 ($2414.68).
- **Internal Audit:**
  - Expenditures are present only in 2015 ($293.79).

And so on, Patterns of growth, decline, and fluctuations in expenditure can be observed across divisions and years.

1.3.3 3. Visualization

- 3.1 Craft a visual representation showcasing the temporal evolution of expenditures by Division.

- 3.2 Focus primarily on transaction amounts but also consider integrating other insightful metrics.

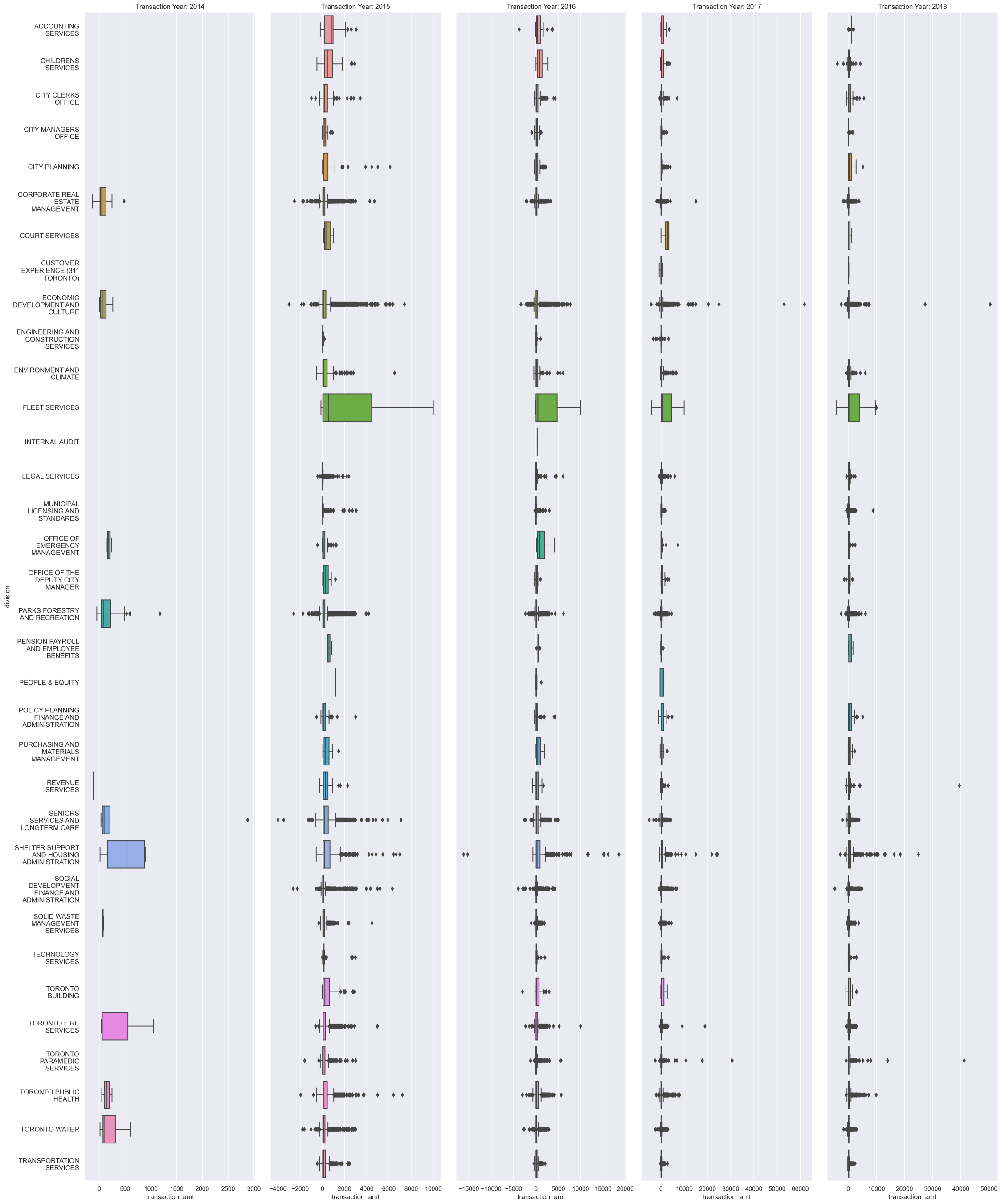
Note: - Combine smaller Divisions for clarity. - Showcase multiple visualizations, if necessary, to portray the data most effectively and reveal significant insights.

3.1 Box Plot Analysis showcasing the temporal evolution of expenditures by Division Y-o-Y.

[ ]:

```
import textwrap

# Get the unique transaction years from the 'transaction_year' column
unique_years = combined_pcard['transaction_year'].unique()
unique_years.sort()

# Get an ordered list of divisions and wrap the text
ordered_divisions = combined_pcard['division'].sort_values().unique()
wrapped_divisions = ['\n'.join(textwrap.wrap(div, width=15)) for div in␣
 ↪ordered_divisions]

# Create a figure and axes objects for subplots
```

```
fig, axes = plt.subplots(nrows=1, ncols=len(unique_years),␣
 ↪figsize=(5*len(unique_years), 30))

# For each year, create a boxplot
for idx, year in enumerate(unique_years):
    yearly_data = combined_pcard[combined_pcard['transaction_year'] == year]
    sns.boxplot(x="transaction_amt", y="division", data=yearly_data,␣
 ↪ax=axes[idx], order=ordered_divisions)
    axes[idx].set_title(f'Transaction Year: {year}')

    # Remove y-axis labels for plots after the first one and adjust font size
    if idx > 0:
        axes[idx].set_yticklabels([])
        axes[idx].set_ylabel('')
    else:
        axes[idx].set_yticklabels(wrapped_divisions, fontsize=12)

plt.tight_layout()
plt.show()
```



[ ]:
```
#3.1.1
#Extract box-plot statistics for further analysis
# Adjust display setting
pd.set_option('display.max_rows', None)
pd.set_option('display.expand_frame_repr', False)
```

```
# List to store the results
results = []

# Unique years and divisions
unique_years = combined_pcard['transaction_year'].unique()
unique_years.sort()
ordered_divisions = combined_pcard['division'].sort_values().unique()

# Extract box plot data for each year and division
for year in unique_years:
    for division in ordered_divisions:
        subset = combined_pcard[(combined_pcard['transaction_year'] == year) &␣
 ↪(combined_pcard['division'] == division)]

        # Compute the required statistics
        min_val = subset['transaction_amt'].min()
        q1 = subset['transaction_amt'].quantile(0.25)
        median = subset['transaction_amt'].median()
        q3 = subset['transaction_amt'].quantile(0.75)
        max_val = subset['transaction_amt'].max()

        results.append([year, division, min_val,q1, median, q3, max_val])

# Convert to a DataFrame
df_boxplot_data = pd.DataFrame(results, columns=['Year', 'Division', 'Minimum␣
 ↪Expenditures by Division', 'Q1', 'Median Expenditures by Division',
                                                  'Q3', 'Maximum Expenditures␣
 ↪by Division'])

# # Pivot the DataFrame with multi-level columns
pivoted = df_boxplot_data.set_index(['Year', 'Division']).unstack('Division')


# Formatting function for currency
def format_currency(val):
    if pd.notna(val):
        try:
            numeric_val = float(val)
            return "{:,.2f}".format(numeric_val)
        except ValueError:
            return str(val)  # If not numeric, return the original value as a␣
 ↪string
    else:
        return ""

# Apply the formatting function across the DataFrame
formatted_pivoted = pivoted.applymap(format_currency)
```

```
formatted_pivoted.T
```

```
[ ]: Year
     2014       2015        2016       2017       2018
                                     Division
     Minimum Expenditures by Division ACCOUNTING SERVICES
     -199.86   -3,744.30     -35.00      90.46
                                     CHILDRENS SERVICES
     -499.00       21.30    -165.00  -3,842.00
                                     CITY CLERKS OFFICE
     -1,000.00     -317.76    -584.25    -475.00
                                     CITY MANAGERS OFFICE
     -22.60     -899.54     -22.60     -22.60
                                     CITY PLANNING
     8.00     -371.52     -14.53       8.00
                                     CORPORATE REAL ESTATE MANAGEMENT
     -131.02  -2,527.71   -2,163.00  -1,836.25  -1,798.27
                                     COURT SERVICES
     103.88                  19.66      31.63
                                     CUSTOMER EXPERIENCE (311 TORONTO)
     -734.50      18.59
                                     ECONOMIC DEVELOPMENT AND CULTURE
     6.10  -3,000.00   -3,338.87  -4,234.54  -2,486.00
                                     ENGINEERING AND CONSTRUCTION SERVICES
     -20.00        1.00  -3,334.26
                                     ENVIRONMENT AND CLIMATE
     -540.00     -446.35    -231.28    -669.53
                                     FLEET SERVICES
     -128.00     -130.86  -3,986.00  -4,249.25
                                     INTERNAL AUDIT
     293.79
                                     LEGAL SERVICES
     -440.70     -141.25    -785.35    -542.40
                                     MUNICIPAL LICENSING AND STANDARDS
     -30.00     -209.94     -35.46    -599.88
                                     OFFICE OF EMERGENCY MANAGEMENT
     135.60    -446.80      112.99       7.33      24.39
                                     OFFICE OF THE DEPUTY CITY MANAGER
     21.77     -395.55     -11.30  -1,400.00
                                     PARKS FORESTRY AND RECREATION
     -46.15  -2,587.70   -2,332.32  -3,000.00  -2,484.00
                                     PENSION PAYROLL AND EMPLOYEE BENEFITS
     450.87      111.87     -89.27      84.75
                                     PEOPLE & EQUITY
     1,199.40      -12.77    -399.80
                                     POLICY PLANNING FINANCE AND ADMINISTRATION
```

```
-524.42     -310.98  -1,017.00      13.57
                                 PURCHASING AND MATERIALS MANAGEMENT
33.88       33.88    -318.66       7.62
                                 REVENUE SERVICES
-112.86    -271.06     -777.02    -177.47    -627.77
                                 SENIORS SERVICES AND LONGTERM CARE
35.64  -4,000.00   -2,547.19  -5,022.49  -2,009.13
                                 SHELTER SUPPORT AND HOUSING ADMINISTRATION
16.03    -560.98  -16,150.00    -511.78  -2,830.00
                                 SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION
-2,639.12   -3,949.35    -847.50  -4,746.00
                                 SOLID WASTE MANAGEMENT SERVICES
57.57    -337.87   -1,050.90    -640.71    -503.91
                                 TECHNOLOGY SERVICES
2.12       28.74     -62.09      -4.99
                                 TORONTO BUILDING
0.55   -2,970.67     -44.09    -852.60
                                 TORONTO FIRE SERVICES
45.62    -622.68   -2,272.69    -395.50    -768.01
                                 TORONTO PARAMEDIC SERVICES
-1,606.46   -1,205.71  -2,500.00    -716.99
                                 TORONTO PUBLIC HEALTH
50.68  -1,950.00   -3,000.00  -2,059.42    -795.00
                                 TORONTO WATER
17.76  -1,790.10   -2,807.00  -2,339.10  -1,578.09
                                 TRANSPORTATION SERVICES
-455.82     -398.85    -292.25    -168.37
Q1                               ACCOUNTING SERVICES
194.79      147.13     151.26   1,107.40
                                 CHILDRENS SERVICES
186.21      367.25     132.46     120.00
                                 CITY CLERKS OFFICE
66.32       57.34      56.74      54.34
                                 CITY MANAGERS OFFICE
60.55       49.88      83.66      34.61
                                 CITY PLANNING
55.51       39.60      36.70      84.18
                                 CORPORATE REAL ESTATE MANAGEMENT
16.54      33.90       38.03      33.81      34.58
                                 COURT SERVICES
208.41               1,784.66      66.50
                                 CUSTOMER EXPERIENCE (311 TORONTO)
-57.22      29.50
                                 ECONOMIC DEVELOPMENT AND CULTURE
26.12      26.77       22.86      26.32      20.69
                                 ENGINEERING AND CONSTRUCTION SERVICES
9.00       30.35      30.55
```

```
                                 ENVIRONMENT AND CLIMATE
37.55       36.76      51.76      53.58
                                 FLEET SERVICES
30.00       45.09      57.00      50.00
                                 INTERNAL AUDIT
293.79
                                 LEGAL SERVICES
3.50       11.50      56.50      95.62
                                 MUNICIPAL LICENSING AND STANDARDS
10.17       11.78      56.98      35.22
                                 OFFICE OF EMERGENCY MANAGEMENT
161.06      20.05      357.01      24.86      39.30
                                 OFFICE OF THE DEPUTY CITY MANAGER
120.26       39.02      77.83      24.20
                                 PARKS FORESTRY AND RECREATION
44.72      33.41       33.90      33.49      31.57
                                 PENSION PAYROLL AND EMPLOYEE BENEFITS
479.12      450.87      36.05      190.69
                                 PEOPLE & EQUITY
1,199.40       20.13    -374.81
                                 POLICY PLANNING FINANCE AND ADMINISTRATION
22.60       22.60      91.12      86.16
                                 PURCHASING AND MATERIALS MANAGEMENT
151.22      141.38      57.90      55.28
                                 REVENUE SERVICES
-112.86      87.27       45.18      39.53      46.39
                                 SENIORS SERVICES AND LONGTERM CARE
60.00      58.46       56.42      49.78      60.24
                                 SHELTER SUPPORT AND HOUSING ADMINISTRATION
160.91      50.85       72.19      69.13      92.64
                                 SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION
19.02       22.41      22.00      22.67
                                 SOLID WASTE MANAGEMENT SERVICES
64.19      38.94       39.73      49.64      43.98
                                 TECHNOLOGY SERVICES
90.65       74.09      61.77      67.39
                                 TORONTO BUILDING
44.99       74.66     152.53      70.09
                                 TORONTO FIRE SERVICES
52.42      50.31       46.33      54.91      56.50
                                 TORONTO PARAMEDIC SERVICES
15.00       16.00      15.00      20.41
                                 TORONTO PUBLIC HEALTH
100.51      45.20       39.25      39.00      36.16
                                 TORONTO WATER
73.81      41.79       41.97      40.64      41.18
                                 TRANSPORTATION SERVICES
```

```
22.58       24.64      23.04      34.30
Median Expenditures by Division  ACCOUNTING SERVICES
804.19      367.22   1,084.80   1,107.40
                                 CHILDRENS SERVICES
447.68      819.25     471.77     187.57
                                 CITY CLERKS OFFICE
186.71      188.06     142.98     223.36
                                 CITY MANAGERS OFFICE
155.12      150.00     101.39      51.78
                                 CITY PLANNING
139.13      113.00     118.65     337.87
                                 CORPORATE REAL ESTATE MANAGEMENT
32.32      79.10       90.32      76.39      86.70
                                 COURT SERVICES
317.04               2,878.68     148.31
                                 CUSTOMER EXPERIENCE (311 TORONTO)
250.00      45.18
                                 ECONOMIC DEVELOPMENT AND CULTURE
57.86      79.10       75.77      87.27      73.43
                                 ENGINEERING AND CONSTRUCTION SERVICES
17.00       32.45      31.60
                                 ENVIRONMENT AND CLIMATE
120.00      158.66     186.61     190.97
                                 FLEET SERVICES
539.00      457.00     695.09     272.00
                                 INTERNAL AUDIT
293.79
                                 LEGAL SERVICES
9.50       60.00     172.50     175.00
                                 MUNICIPAL LICENSING AND STANDARDS
14.76       29.97     100.00      95.00
                                 OFFICE OF EMERGENCY MANAGEMENT
186.53      72.52      819.79     136.49     106.62
                                 OFFICE OF THE DEPUTY CITY MANAGER
250.30      151.82     173.74      90.85
                                 PARKS FORESTRY AND RECREATION
79.08      82.50       83.83      82.83      75.11
                                 PENSION PAYROLL AND EMPLOYEE BENEFITS
507.37      502.88      70.61     629.98
                                 PEOPLE & EQUITY
1,199.40       51.69     504.38
                                 POLICY PLANNING FINANCE AND ADMINISTRATION
117.29      135.60     294.96     470.85
                                 PURCHASING AND MATERIALS MANAGEMENT
293.82      422.77     188.12     347.52
                                 REVENUE SERVICES
-112.86     248.60      111.33     132.23     121.38
```

```
                                 SENIORS SERVICES AND LONGTERM CARE
94.29     174.80      164.30     135.54     152.94
                                 SHELTER SUPPORT AND HOUSING ADMINISTRATION
539.62     203.09      279.11     255.38     319.76
                                 SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION
44.29       52.98      54.84      52.67
                                 SOLID WASTE MANAGEMENT SERVICES
70.81      84.80       80.34      95.00     100.00
                                 TECHNOLOGY SERVICES
110.25      136.70     143.18     168.57
                                 TORONTO BUILDING
214.07      199.44     440.00     153.69
                                 TORONTO FIRE SERVICES
59.23      73.45       65.19      90.40      97.11
                                 TORONTO PARAMEDIC SERVICES
57.82       57.00      65.87      87.02
                                 TORONTO PUBLIC HEALTH
150.34     134.61      150.20     145.00     150.22
                                 TORONTO WATER
96.14     100.52      108.55     103.87     103.30
                                 TRANSPORTATION SERVICES
101.69      101.70      96.03      95.31
Q3                               ACCOUNTING SERVICES
961.90    1,084.80   1,084.80   1,107.40
                                 CHILDRENS SERVICES
896.87    1,400.00   1,024.34     450.94
                                 CITY CLERKS OFFICE
446.35      452.00     448.91     846.55
                                 CITY MANAGERS OFFICE
305.25      378.64     279.75      75.50
                                 CITY PLANNING
499.57      438.44     296.53   1,200.86
                                 CORPORATE REAL ESTATE MANAGEMENT
132.02     213.59      245.33     214.30     243.99
                                 COURT SERVICES
737.98               3,384.35     632.79
                                 CUSTOMER EXPERIENCE (311 TORONTO)
495.97     186.19
                                 ECONOMIC DEVELOPMENT AND CULTURE
134.71     312.20      309.95     363.24     264.38
                                 ENGINEERING AND CONSTRUCTION SERVICES
45.00      119.45      93.22
                                 ENVIRONMENT AND CLIMATE
426.65      426.16     455.99     478.79
                                 FLEET SERVICES
4,440.00    4,746.00   4,624.81   3,936.00
                                 INTERNAL AUDIT
```

```
293.79
                                 LEGAL SERVICES
27.00      169.50     325.00     381.38
                                 MUNICIPAL LICENSING AND STANDARDS
41.14      100.00     253.31     250.00
                                 OFFICE OF EMERGENCY MANAGEMENT
212.00     223.20    1,991.52     341.69     323.86
                                 OFFICE OF THE DEPUTY CITY MANAGER
521.11      333.78     819.25     399.10
                                 PARKS FORESTRY AND RECREATION
226.00     216.43      225.72     222.82     198.32
                                 PENSION PAYROLL AND EMPLOYEE BENEFITS
677.43      551.36     231.58   1,199.21
                                 PEOPLE & EQUITY
1,199.40      143.08   1,180.64
                                 POLICY PLANNING FINANCE AND ADMINISTRATION
254.16      310.98   1,084.80   1,107.40
                                 PURCHASING AND MATERIALS MANAGEMENT
593.99    1,028.40     573.77     739.79
                                 REVENUE SERVICES
-112.86     498.72      610.20     354.84     426.25
                                 SENIORS SERVICES AND LONGTERM CARE
209.61     523.07      483.95     366.30     391.58
                                 SHELTER SUPPORT AND HOUSING ADMINISTRATION
878.14     678.68      922.08     794.67     792.90
                                 SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION
123.97      145.94     156.45     150.38
                                 SOLID WASTE MANAGEMENT SERVICES
77.43     180.19      179.67     225.99     250.36
                                 TECHNOLOGY SERVICES
145.66      187.96     270.52     309.25
                                 TORONTO BUILDING
638.02      727.61   1,275.76     904.88
                                 TORONTO FIRE SERVICES
557.15     276.85      290.33     283.79     352.91
                                 TORONTO PARAMEDIC SERVICES
224.44      130.00     219.76     291.85
                                 TORONTO PUBLIC HEALTH
200.17     438.15      502.14     432.40     425.00
                                 TORONTO WATER
315.53     226.04      227.47     233.15     248.59
                                 TRANSPORTATION SERVICES
268.93      268.05     223.74     225.99
Maximum Expenditures by Division ACCOUNTING SERVICES
3,048.46    3,744.30   3,610.35   1,929.99
                                 CHILDRENS SERVICES
2,909.75    2,720.00   3,904.15   4,294.00
```

```
                                 CITY CLERKS OFFICE
3,424.15    4,334.01   6,942.70   5,554.02
                                 CITY MANAGERS OFFICE
908.42    1,207.39   2,541.79   1,704.66
                                 CITY PLANNING
6,107.56    2,226.00   4,044.65   5,256.79
                                 CORPORATE REAL ESTATE MANAGEMENT
482.40   4,677.30    3,312.60  15,010.37   3,842.00
                                 COURT SERVICES
1,000.00               3,384.35   1,107.40
                                 CUSTOMER EXPERIENCE (311 TORONTO)
827.14     243.18
                                 ECONOMIC DEVELOPMENT AND CULTURE
265.59   7,400.00    7,694.42  61,842.37  50,413.79
                                 ENGINEERING AND CONSTRUCTION SERVICES
198.10    1,054.64   3,334.26
                                 ENVIRONMENT AND CLIMATE
6,515.02    6,075.78   6,667.00   6,073.75
                                 FLEET SERVICES
10,000.00   10,000.00  10,000.00  10,000.00
                                 INTERNAL AUDIT
293.79
                                 LEGAL SERVICES
2,395.60    6,068.52   5,987.18   2,521.30
                                 MUNICIPAL LICENSING AND STANDARDS
3,046.51    2,997.89   2,034.46   8,859.20
                                 OFFICE OF EMERGENCY MANAGEMENT
237.46   1,265.60    4,187.78   7,289.63   2,509.11
                                 OFFICE OF THE DEPUTY CITY MANAGER
1,173.55    1,030.00   3,467.84   1,600.00
                                 PARKS FORESTRY AND RECREATION
1,180.19   4,186.65    6,158.75   4,622.95   6,162.12
                                 PENSION PAYROLL AND EMPLOYEE BENEFITS
847.50      932.25   1,015.87   1,695.00
                                 PEOPLE & EQUITY
1,199.40    1,199.40   1,199.40
                                 POLICY PLANNING FINANCE AND ADMINISTRATION
3,000.00    4,294.00   4,746.90   5,210.00
                                 PURCHASING AND MATERIALS MANAGEMENT
1,467.87    1,919.20   2,687.12   2,284.86
                                 REVENUE SERVICES
-112.86   2,282.60    1,717.23   3,164.00  39,493.50
                                 SENIORS SERVICES AND LONGTERM CARE
2,881.50   7,079.45    4,969.51   4,282.70   3,836.35
                                 SHELTER SUPPORT AND HOUSING ADMINISTRATION
902.48   6,999.28   18,573.46  24,343.75  25,000.00
                                 SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION
```

```
6,305.00    4,237.50   6,780.00   4,853.35
                                 SOLID WASTE MANAGEMENT SERVICES
84.05   4,469.57    1,937.30   4,514.35   3,692.78
                                 TECHNOLOGY SERVICES
2,966.16    2,017.05   3,164.00   2,904.89
                                 TORONTO BUILDING
2,919.58    2,994.50   2,745.90   2,998.00
                                 TORONTO FIRE SERVICES
1,055.08   4,968.00   10,000.00  18,958.50   2,999.44
                                 TORONTO PARAMEDIC SERVICES
3,000.00    5,619.00  30,650.50  41,205.75
                                 TORONTO PUBLIC HEALTH
250.00   7,211.75    5,671.38   8,085.15   9,945.70
                                 TORONTO WATER
603.88   3,000.00    2,928.09   2,959.72   2,932.35
                                 TRANSPORTATION SERVICES
2,467.02    2,092.97   2,841.09   2,431.66
```

```
[ ]: # combined_pcard[(combined_pcard['division'] == 'SHELTER SUPPORT AND HOUSING␣
     ↪ADMINISTRATION') & (combined_pcard['transaction_year'] == 2016)].
     ↪sort_values(by='transaction_amt')
```

3.1.2 Interpratation of the Box plot statistics

- The data shows a dynamic financial landscape with various divisions undergoing changes in expenditures Year on Year. Some divisions show consistent trends(E.g `CITY MANAGER'S OFFICE`), while others fluctuate significantly.(E.g. `ACCOUNTING SERVICES`)
- **Comparative Expenditure:** Divisions such as `TECHNOLOGY SERVICES` and `TORONTO WATER` have similar ranges of expenditures across the years.
- **Missing Data** Some divisions have missing data for certain years, such as `CUSTOMER EXPERIENCE (311 TORONTO)`. It's important to understand if these divisions were non-operational in those years or if data is simply unavailable.
- **Division Variability** Some divisions have relatively stable expenditures over the years, such as `FLEET SERVICES`. This consistency might indicate a capped expenditure or a budget that's rigorously adhered to. In contrast, other divisions like `TORONTO PARAMEDIC SERVICES` show significant fluctuation, with the maximum expenditure jumping from $3,000 to over $41,000 within the period. Also, `ECONOMIC DEVELOPMENT AND CULTURE`, show a dramatic increase in their maximum expenditures, from a mere $265 in one year to a whopping $61,842 in another.
- **Potential Anomalies** Many divisions have negative expenditures, understanding the exact implications would require more context on the G/L accounts and Purpose of purchase. E.g. The `SHELTER SUPPORT AND HOUSING ADMINISTRATION` division shows significant negative expenditures in 2016 of -$16,150 and -$15,246 as a result of refunds for gifts from clients during LD. These negative values requires further investigation with the government on such transactions. On the other end `SHELTER SUPPORT AND HOUSING ADMINISTRATION` has been

increasing their maximum expenditures year over year, potentially indicating an expanding scope of services or higher costs among other divisions.

- While some divisions have escalating costs, others are maintaining or reducing their maximum expenditures. The reasons for these could be multifaceted, ranging from inflation, increased operational scope, cost-saving measures, economic events affecting the city, or changes in budget allocations.

A deep dive of the specific causes for these expenditure trends, can reveal the factors at play. For a more in-depth understanding and actionable insights, it would be beneficial to collaborate with a financial analyst familiar with the city's budgeting and finance operations or interrogate the data further for more insights.

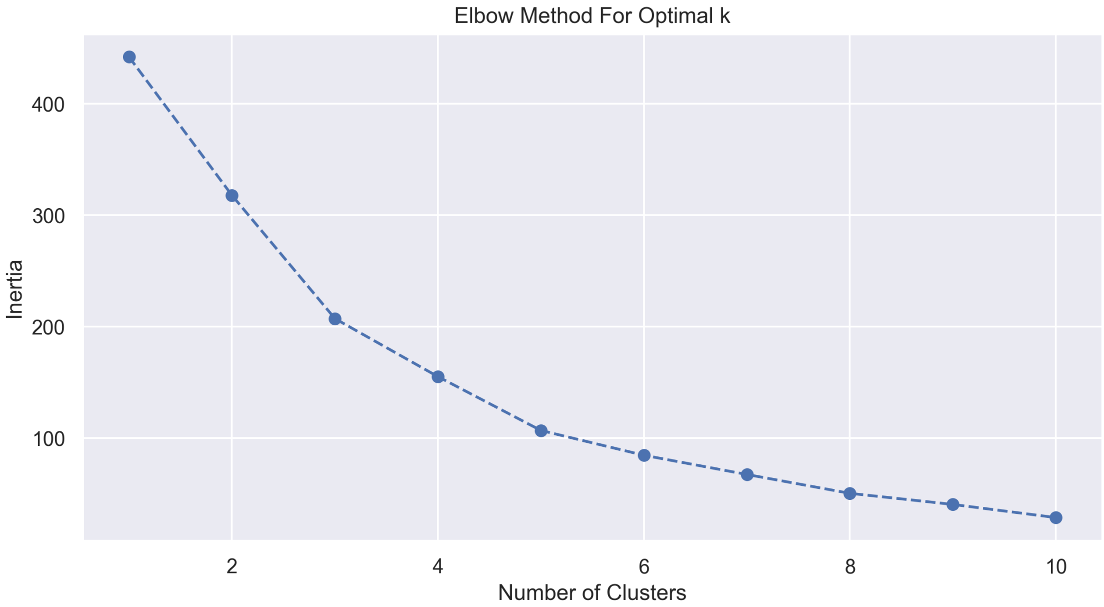
1.3.4 4. Cluster Creation

4.1 Group Divisions with similar behaviours based on the profiles developed in the prior step.

- For clustering we will create new features for our baseline model. We will use KMeans clustering algorithim with silohutte score for measuring the optimal number of clusters we should be having for our new dataframe(`division_features`).
- Our new data set will comprise of
  - `division`
  - unique counts of `batch_transaction_id`,`gl_accounts`,`merchant_name` and `purpose`
  - mean,sum, standard deviation, minium, maximum, median & IQR of the `transaction amounts`
- This will help us have a better understanding of the data before we can recommend an improved model.

[ ]:

```
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Create features for each division
division_features = combined_pcard.groupby(['division']).agg({
    'batch_transaction_id':'nunique',
    'transaction_amt': ['mean', 'sum', 'std', 'min', 'max', 'median',
                        lambda x: x.quantile(0.25), lambda x: x.quantile(0.75)],
    'gl_account': 'nunique',
    'merchant_name': 'nunique',
    'purpose':'nunique'
}).reset_index()

# Update the columns names after aggregation
division_features.columns = ["_".join(x) for x in division_features.columns.
 ↪ravel()]
division_features = division_features.rename(columns={'division_': 'division',
                                                ␣
 ↪'transaction_amt_<lambda_0>': 'transaction_amt_q1',
                                                ␣
 ↪'transaction_amt_<lambda_1>': 'transaction_amt_q3'})

# Compute the IQR
division_features['transaction_amt_iqr'] =␣
 ↪division_features['transaction_amt_q3'] -␣
 ↪division_features['transaction_amt_q1']

# Handling NaN values
# We are replacing NaN values with the mean of the respective columns
for col in division_features.columns:
    if division_features[col].dtype != 'object':  # Avoid filling for␣
 ↪'division' column
        division_features[col].fillna(division_features[col].mean(),␣
 ↪inplace=True)

# Scale the features (excluding 'division')
scaler = StandardScaler()
scaled_features = scaler.fit_transform(division_features.drop(['division'],␣
 ↪axis=1))

# Elbow method to find optimal number of clusters
inertia = []
for i in range(1, 11):  # Testing for up to 10 clusters
    kmeans = KMeans(n_clusters=i, random_state=0)
    kmeans.fit(scaled_features)
    inertia.append(kmeans.inertia_)

# Plotting the elbow curve
plt.figure(figsize=(10, 5))
plt.plot(range(1, 11), inertia, marker='o', linestyle='--')
plt.title('Elbow Method For Optimal k')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia')
plt.show()
```



[ ]:

```
from sklearn.metrics import silhouette_score

# Define the range of clusters
cluster_range = range(4, 15)

# List to store silhouette scores
silhouette_scores = []

# Calculate silhouette scores for each number of clusters
for n_clusters in cluster_range:
    kmeans = KMeans(n_clusters=n_clusters, random_state=42)
    cluster_assignments = kmeans.fit_predict(scaled_features)
    silhouette_avg = silhouette_score(scaled_features, cluster_assignments)
    silhouette_scores.append(silhouette_avg)

# Create a DataFrame to store the results
df_silhouette = pd.DataFrame({
    'Number of Clusters': cluster_range,
    'Silhouette Score': silhouette_scores
})

# Display the DataFrame
df_silhouette

def highlight_max(s):
    """
    Highlight the maximum in a Series yellow.
    """
    is_max = s == s.max()
    return ['background-color: yellow' if v else '' for v in is_max]

# Apply the styling function to the DataFrame
styled_df = df_silhouette.style.apply(highlight_max, subset=['Silhouette␣
 ↪Score'])

# Display the styled DataFrame
styled_df
```

[ ]: <pandas.io.formats.style.Styler at 0x7f9b029e56c0>

[ ]:

```
# Based on the above plot results, we can see that the mean silhouette␣
 ↪Coefficient is highest at 5 clusters (0.3630),
# indicating improved model performance and better cluster assignment for each␣
 ↪dataset instance.

kmeans = KMeans(n_clusters=5,random_state = 42)
division_features['cluster'] = kmeans.fit_predict(scaled_features)

division_features.sort_values(by='cluster').applymap(format_currency)
```

[ ]:

```
                                         division batch_transaction_id_nunique
transaction_amt_mean transaction_amt_sum transaction_amt_std transaction_amt_min
transaction_amt_max transaction_amt_median transaction_amt_q1 transaction_amt_q3
gl_account_nunique merchant_name_nunique purpose_nunique transaction_amt_iqr
cluster
30                     TORONTO PARAMEDIC SERVICES                     2,074.00
295.32          612,783.56            1,371.50           -2,500.00
41,205.75                  62.28              16.00              210.76
77.00                568.00        1,032.00              194.76    0.00
8                ECONOMIC DEVELOPMENT AND CULTURE                    14,846.00
363.97        5,403,507.87            1,164.00           -4,234.54
61,842.37                  79.63              24.85              316.40
147.00              4,174.00        8,704.00              291.55    0.00
24     SHELTER SUPPORT AND HOUSING ADMINISTRATION                     3,085.00
805.01        2,483,463.91            2,090.86          -16,150.00
25,000.00                 265.21              66.67              745.39
131.00                749.00        2,546.00              678.72    0.00
22                               REVENUE SERVICES                       240.00
502.82          120,676.44            2,593.37             -777.02
39,493.50                 131.06              48.79              408.30
40.00                 98.00          213.00              359.51    0.00
17                  PARKS FORESTRY AND RECREATION                    98,337.00
199.38       19,606,012.25              341.98           -3,000.00
6,162.12                  81.81              33.48             218.47
```

```
219.00              5,743.00       55,494.00              184.99    1.00
16              OFFICE OF THE DEPUTY CITY MANAGER                       114.00
328.11           37,404.59              570.55           -1,400.00
3,467.84                 143.69              29.06             446.78
28.00                 76.00          106.00              417.72    2.00
31                          TORONTO PUBLIC HEALTH                     4,936.00
387.27        1,911,541.00              733.21           -3,000.00
9,945.70                 145.00              39.50             447.48
106.00              1,619.00          975.00              407.98    2.00
29                          TORONTO FIRE SERVICES                     4,319.00
293.61        1,268,108.36              613.57           -2,272.69
18,958.50                  77.20              50.59             282.79
95.00                959.00        2,280.00              232.20    2.00
28                               TORONTO BUILDING                       188.00
613.27          115,294.34              857.18           -2,970.67
2,998.00                 247.69              68.47             852.60
28.00                 66.00          155.00              784.13    2.00
27                            TECHNOLOGY SERVICES                       241.00
312.73           75,367.47              601.50              -62.09
3,164.00                 136.70              74.09             266.58
18.00                 86.00          172.00              192.49    2.00
26                SOLID WASTE MANAGEMENT SERVICES                     3,190.00
205.59          655,826.73              355.12           -1,050.90
4,514.35                  90.39              42.94             211.87
125.00                523.00        2,321.00              168.93    2.00
25  SOCIAL DEVELOPMENT FINANCE AND ADMINISTRATION                     5,458.00
235.06        1,282,970.63              620.60           -4,746.00
6,780.00                  50.84              21.45             144.99
64.00              1,153.00        4,206.00              123.54    2.00
23             SENIORS SERVICES AND LONGTERM CARE                     5,298.00
396.27        2,099,412.19              654.44           -5,022.49
7,079.45                 156.50              54.56             449.65
83.00              1,034.00        3,677.00              395.09    2.00
21            PURCHASING AND MATERIALS MANAGEMENT                       110.00
448.65           49,351.13              502.46             -318.66
2,687.12                 286.17              71.67             727.17
23.00                 63.00          105.00              655.50    2.00
20     POLICY PLANNING FINANCE AND ADMINISTRATION                       336.00
500.14          168,046.99              757.78           -1,017.00
5,210.00                 210.87              47.04             820.31
24.00                116.00          144.00              773.27    2.00
19                                PEOPLE & EQUITY                       112.00
201.52           22,570.71              386.83             -399.80
1,199.40                  65.55              20.04             179.14
4.00                 60.00           13.00              159.09    2.00
33                        TRANSPORTATION SERVICES                     1,397.00
214.00          298,957.10              341.93             -455.82
```

```
2,841.09                  98.50              25.09             242.90
101.00                330.00          910.00              217.81    2.00
15                 OFFICE OF EMERGENCY MANAGEMENT                       151.00
372.25           56,209.00              835.95             -446.80
7,289.63                  92.94              30.00             337.85
41.00                 94.00          111.00              307.85    2.00
2                              CITY CLERKS OFFICE                       587.00
456.69          268,076.01              769.35           -1,000.00
6,942.70                 187.45              59.76             538.48
49.00                294.00          475.00              478.72    2.00
3                            CITY MANAGERS OFFICE                       455.00
175.43           79,819.24              292.87             -899.54
2,541.79                  76.32              39.00             198.18
21.00                331.00          250.00              159.18    2.00
4                                   CITY PLANNING                       350.00
508.32          177,912.29              875.40             -371.52
6,107.56                 129.94              46.45             552.62
17.00                 97.00          278.00              506.18    2.00
5                CORPORATE REAL ESTATE MANAGEMENT                    14,700.00
199.47        2,935,159.62              357.30           -2,527.71
15,010.37                  83.87              35.00             227.99
99.00              1,179.00        6,702.00              192.99    2.00
32                                  TORONTO WATER                    11,709.00
203.36        2,381,143.08              334.34           -2,807.00
3,000.00                 104.37              41.32             230.57
138.00              1,334.00        7,046.00              189.25    2.00
9           ENGINEERING AND CONSTRUCTION SERVICES                     1,426.00
55.53           79,184.69              217.06           -3,334.26
3,334.26                  30.55              10.00              63.90
9.00                 19.00          604.00               53.90    2.00
7               CUSTOMER EXPERIENCE (311 TORONTO)                        20.00
191.32            3,826.44              363.13             -734.50
827.14                 169.34              25.63             313.75
6.00                 13.00           20.00              288.11    2.00
13                                 LEGAL SERVICES                     5,821.00
93.71          545,494.58              251.78             -785.35
6,068.52                  17.00               6.50             100.00
18.00                116.00          210.00               93.50    2.00
14              MUNICIPAL LICENSING AND STANDARDS                       904.00
193.70          175,101.84              480.15             -599.88
8,859.20                  50.00              13.87             167.14
77.00                270.00          403.00              153.27    2.00
10                        ENVIRONMENT AND CLIMATE                       701.00
442.94          310,498.49              875.30             -669.53
6,667.00                 169.50              44.35             454.88
45.00                335.00          429.00              410.53    2.00
11                                 FLEET SERVICES                     2,005.00
```

```
2,711.51        5,436,574.33            3,693.09           -4,249.25
10,000.00                 491.00              55.00           4,543.25
37.00                97.00          388.00            4,488.25    3.00
6                                  COURT SERVICES                        18.00
842.28           15,160.97            1,102.37               19.66
3,384.35                 263.05             114.99           1,080.55
10.00                14.00           17.00              965.56    4.00
12                                 INTERNAL AUDIT                         1.00
293.79              293.79              818.42              293.79
293.79                 293.79             293.79             293.79
1.00                  1.00            1.00                0.00    4.00
18          PENSION PAYROLL AND EMPLOYEE BENEFITS                        28.00
447.04           12,517.14              420.37              -89.27
1,695.00                 450.87              88.14             625.39
7.00                  9.00           20.00              537.25    4.00
1                              CHILDRENS SERVICES                       345.00
651.46          224,753.04              852.27           -3,842.00
4,294.00                 367.82             140.00             853.15
13.00               130.00          180.00              713.15    4.00
0                             ACCOUNTING SERVICES                       218.00
732.83          159,757.66              730.37           -3,744.30
3,744.30                 804.19             194.79           1,084.80
22.00                59.00          126.00              890.01    4.00
```

4.2 Describe each cluster and elucidate its unique attributes.

```
[ ]: clusters = division_features['cluster'].nunique()
cluster_descriptions = []

for cluster in range(clusters):
    cluster_data = division_features[division_features['cluster'] == cluster]

    # Fetching basic stats
    mean_transaction_amt = cluster_data['transaction_amt_mean'].mean()
    median_transaction_amt = cluster_data['transaction_amt_median'].mean()
    max_transaction_amt = cluster_data['transaction_amt_max'].max()
    min_transaction_amt = cluster_data['transaction_amt_min'].min()
    num_divisions = cluster_data['division'].nunique()

    description = {
        "Cluster": cluster,
        "Num_Divisions": num_divisions,
        "Avg_Mean_Transaction": mean_transaction_amt,
        "Avg_Median_Transaction": median_transaction_amt,
        "Max_Transaction": max_transaction_amt,
        "Min_Transaction": min_transaction_amt
    }
    cluster_descriptions.append(description)

# Convert list of dictionaries to DataFrame
description_df = pd.DataFrame(cluster_descriptions)

description_df.applymap(format_currency)
```

[ ]:

```
  Cluster Num_Divisions Avg_Mean_Transaction Avg_Median_Transaction
Max_Transaction Min_Transaction
0    0.00          4.00               491.78                 134.54
61,842.37      -16,150.00
1    1.00          1.00               199.38                  81.81
6,162.12       -3,000.00
2    2.00         23.00               305.60                 122.63
18,958.50       -5,022.49
3    3.00          1.00             2,711.51                 491.00
10,000.00       -4,249.25
4    4.00          5.00               593.48                 435.94
4,294.00       -3,842.00
```

Cluster 0 (TORONTO PARAMEDIC SERVICES,ECONOMIC DEVELOPMENT AND CULTURE,SHELTER SUPPORT AND HOUSING ADMINISTRATION,REVENUE SERVICES): - Divisions: This cluster comprises 4 divisions. - **Transaction Attributes:** Divisions in this cluster have an average mean transaction of approximately $491.78 and a median transaction of about $134.54. The range of transaction amounts in this cluster is quite wide, spanning from -$16,150.00 to a substantial $61,842.37. - **Unique Attribute:** Cluster 0 represents divisions with diverse transaction behaviors. These divisions handle a wide range of transaction values, possibly indicating occasional high-value transactions.

Cluster 1(PARKS FORESTRY AND RECREATION): - **Divisions:** This cluster is unique as it contains only 1 division. - **Transaction Attributes:** The mean transaction for this division is around $199.38, with a median transaction of $81.81. Transactions in this division range from -$3,000.00 to $6,162.12. - **Unique Attribute:** Despite having only one division, this cluster stands out due to the division's modest transaction values. However, the narrower range of transactions compared to other clusters is notable.

Cluster 2 (OTHERS): - **Divisions:** The largest cluster, containing 23 divisions. - **Transaction Attributes:** On average, divisions in this cluster have a mean transaction amount of about $305.60, with a median transaction of $122.63. Transaction values within this cluster range from -$5,022.49 to $18,958.50. - **Unique Attribute:** Cluster 2, despite having the most divisions, deals with a wide variety of transactions but doesn't have the broadest range of transaction values among the clusters.

Cluster 3(FLEET SERVICES): - **Divisions:** Similar to Cluster 1, Cluster 3 has only 1 division. - **Transaction Attributes:** The mean transaction for this division is notably high at approximately $2,711.51, with a median transaction of $491.00. Transactions here span from -$4,249.25 to $10,000.00. - **Unique Attribute:** With a solitary division, Cluster 3's standout characteristic is its high mean transaction value, suggesting that this specific division handles larger monetary transactions.

Cluster 4(COURT SERVICES,INTERNAL AUDIT,PENSION PAYROLL AND EMPLOYEE BENEFITS,CHILDRENS SERVICES,ACCOUNTING SERVICES): - **Divisions:** This cluster comprises 5 divisions. - **Transaction Attributes:** Divisions in Cluster 4 have an average mean transaction amount of around $593.48. The median transaction for these divisions is approximately $435.94. Transactions within this cluster range from a minimum of -$3,842.00 to a maximum of $4,294.00. - **Unique Attribute:** Cluster 4 showcases divisions with moderate transaction values, representing typical and routine transactions, given the relatively narrow range between the maximum and minimum transaction amounts.

In summary, the analysis of the updated dataset reveals varying transaction behaviors across the different clusters. Clusters 0 and 4, each consisting of multiple divisions, represent more common transaction patterns. Cluster 1 exhibits a broader range of transaction values, while Clusters 2 and 3 highlight unique transaction behaviors of individual divisions.

1.3.5 5. Anomaly Detection

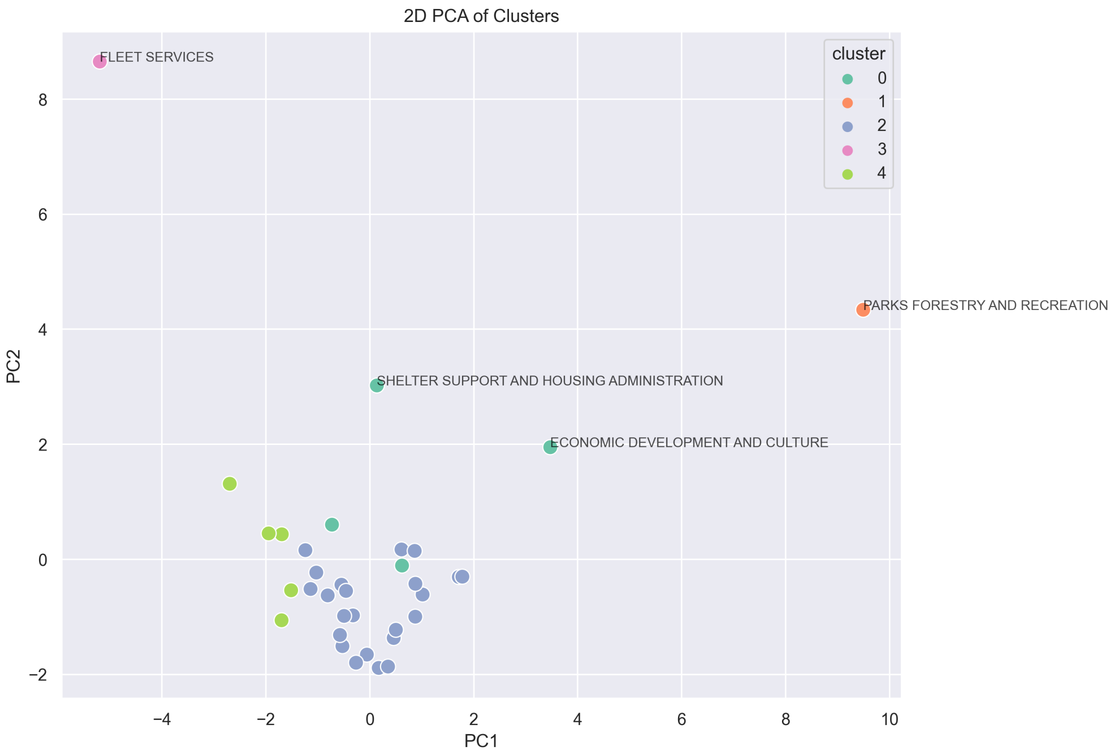
- 5.1 Utilizing the profiles, pinpoint Divisions that manifest anomalous behaviours.
- 5.2 Further subdivide by clusters to highlight any irregularities if beneficial.

[ ]:

```python
# To identify anomalous behavior in divisions based on the provided data, we'll follow a step-by-step approach:

# 1. **Data Standardization**: Since different features have different scales, it's essential to standardize the data to ensure equal weightage to all features.

# 2. **PCA for Visualization**: Principal Component Analysis (PCA) can be used to reduce the dimensionality of the data and visualize it in 2D or 3D space to spot any potential outliers or anomalous behavior.

# 3. **IQR Method**: We can use the Interquartile Range (IQR) method to identify outliers in the principal components. The IQR is a measure of statistical spread between the 25th percentile (Q1) and the 75th percentile (Q3). Anything outside the range `[Q1 - 1.5*IQR, Q3 + 1.5*IQR]` can be considered as an anomaly or outlier.

# 4. **Analysis and Interpretation**: Based on the results, we can then identify divisions that show anomalous behavior.

# Let's walk through the process:

from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
# Step 1: Data Standardization
features = division_features.drop(['division', 'cluster'], axis=1)
scaled_features = StandardScaler().fit_transform(features)

# Step 2: PCA for Visualization
pca = PCA(n_components=2)
principal_components = pca.fit_transform(scaled_features)
principal_df = pd.DataFrame(data=principal_components, columns=['PC1', 'PC2'])
principal_df['division'] = division_features['division']

# Step 3: IQR Method for anomaly detection
Q1 = principal_df[['PC1', 'PC2']].quantile(0.25)
Q3 = principal_df[['PC1', 'PC2']].quantile(0.75)
IQR = Q3 - Q1
filter = (principal_df[['PC1', 'PC2']] < (Q1 - 1.5 * IQR)) | (principal_df[['PC1', 'PC2']] > (Q3 + 1.5 * IQR))

anomalous_divisions_df = principal_df[filter.any(axis=1)]

# Display anomalous divisions
print(anomalous_divisions_df)

# Attach cluster labels from KMeans
principal_df['cluster'] = division_features['cluster']

# Plotting the 2D PCA
plt.figure(figsize=(10, 8))
sns.scatterplot(x='PC1', y='PC2', hue='cluster', data=principal_df, palette='Set2', s=100)

# Annotate each anomalous division with its name
for _, row in anomalous_divisions_df.iterrows():
    plt.annotate(row['division'], (row['PC1'], row['PC2']), fontsize=9, alpha=0.7, color='black')

plt.title('2D PCA of Clusters')
plt.show()
```

```
         PC1       PC2                                  division
8   3.472690  1.950503          ECONOMIC DEVELOPMENT AND CULTURE
11 -5.195409  8.652914                            FLEET SERVICES
17  9.493510  4.338936               PARKS FORESTRY AND RECREATION
24  0.135498  3.023488  SHELTER SUPPORT AND HOUSING ADMINISTRATION
```



5.3 Rank Divisions by their "abnormality", opting for criteria like the Top 10 Divisions or the Top 10% based on specific metrics. We considered `Distance from the Median` approach due to it's simplicity of use:

1. **Distance from the Median:** One common method to rank the abnormality of divisions is to calculate the distance of each point from the median (or mean) of the PCA components. The idea is that the farther a division is from the central tendency of the data, the more abnormal it is.

[ ]:
```
# Calculate the median for PC1 and PC2
median_PC1 = principal_df['PC1'].median()
median_PC2 = principal_df['PC2'].median()

# Compute the distance of each division from the median of the PCA components
anomalous_divisions_df['distance_from_median'] = np.
 ↪sqrt((anomalous_divisions_df['PC1'] - median_PC1)**2 +
                                                       ␣
 ↪(anomalous_divisions_df['PC2'] - median_PC2)**2).round(2)

# Rank the divisions based on the distance
anomalous_divisions_df = anomalous_divisions_df.
 ↪sort_values(by='distance_from_median', ascending=False)
```

```python
# Reset index, drop old index, and start counting from 1
anomalous_divisions_df = anomalous_divisions_df[['division',␣
 ↪'distance_from_median']].reset_index(drop=True)
anomalous_divisions_df.index = anomalous_divisions_df.index + 1

# Display the ranked anomalous divisions
anomalous_divisions_df
```

```
[ ]:                                    division  distance_from_median
     1             PARKS FORESTRY AND RECREATION                 10.91
     2                            FLEET SERVICES                 10.36
     3            ECONOMIC DEVELOPMENT AND CULTURE                4.48
     4  SHELTER SUPPORT AND HOUSING ADMINISTRATION                3.53
```

5.4 Recommend a strategy to differentiate between "normal" and "abnormal" Divisions through sampling. Using sampling to differentiate between normal and abnormal divisions requires a statistical approach that involves examining the distribution of data and determining whether a division's characteristics fall within expected ranges. Here's a strategy we can follow after we have done procdure 1 to 5 above:

1. **Define `Normal` and `Abnormal` Criteria:**
   - We need to decide on criteria that define what is considered `normal` and `abnormal` for divisions based on the specific attributes we're analyzing (e.g., transaction amounts).
   - This could involve setting thresholds based on historical data.
2. **Sampling:**
   - Select a sample of divisions from our dataset. The sample size will depend on the size of our dataset and the desired level of confidence.
   - we can use random sampling or other sampling methods based on our analysis goals.
3. **Compare with Criteria:**
   - Compare the sample statistics with the predefined criteria for `normal` and `abnormal` divisions.
   - Divisions that fall within the expected ranges are considered `normal`, while those that significantly deviate from the criteria may be `abnormal`.
4. **Statistical Tests:**
   - Perform statistical tests if necessary to quantify the differences between sample statistics and criteria.
   - For example, we can use Z-tests, T-tests, or other relevant tests to compare means or other attributes.
5. **Visualization:**
   - Visualize the sample data and their distributions to see how they align with our `normal` and `abnormal` criteria.
6. **Decision Making:**
   - Based on our analysis, make decisions about whether each division is considered `normal` or `abnormal` according to the predefined criteria.
   - Keep in mind that `abnormal` divisions might need further investigation to understand the underlying reasons for deviations.
7. **Iterative Process:**

- This process may require multiple iterations as we refine our criteria, sample size, or analysis methods.

Note: What's considered `normal` and `abnormal` can depend on the context and business requirements. Using statistical techniques and sampling can help us identify divisions that exhibit unusual behaviors or characteristics, leading to a more data-driven decision-making process.