RSM312 Assignment 2 Report

Group 3 (Sailing Ni, David Bianchin, Vivien Abara Ouko)

Task 2

Variety – different forms and formats of the data

The datasets contain P-Card transactions from fiscal years 2014 and 2015 with a small variety in data format, mainly consisting of text fields with dates and monetary values. The variety in format allows for examining different relationships, such as split purchase violations, but we need consistent format for corresponding fields across the two datasets. For example with the *amount* field we will apply the astype() and replace() function to ensure that all rows in that field are converted to a similar float format (float is used rather than integer due to decimal points), whilst the to_datetime() function is used for the *transaction date* and *posted date* fields to ensure that the format and presentation of the dates are consistent.

Labels are identical in the two datasets, which makes it easier for us to conduct comparable analysis. Datasets are delimited by commas, which is prevalent in *MCC*. Since values with commas in MCC are surrounded by double quotes, importing datasets does not cause problems. To further clean the data in convenience of analysis, we will we remove whitespaces in labels, use the strip() function to remove any white spaces surrounding the text, the lower() function to lowercase *MCC* values.

Velocity – frequency of incoming data that needs processing.

The data used is historical as it consists of information that falls between 2013-2014 and since the case did not give a date, we assume that the analysis is being undertaken between 2015-2023. Auditing the data should be done tri-monthly in order to quickly notice individuals whose spending practices appear fraudulent or exceed limits put in place. If OSU continues to analyse yearly only, then individuals who are violating cardholder rules could make the university lose a lot of money. Therefore, this analysis should be very automated as most of the restrictions can be easily coded such as credit limit and transaction limits, allowing the program to be constantly up-to-date and sending out email/warnings to individuals who exceed the limit or are close to exceeding it.

Veracity – trustworthiness of the data

There are missing values for *transaction* and *posted date*, which transactions we ignored in analysis. This reduces the veracity of the data, since we cannot determine when did those transactions take place, if transactions were completed, whether limits were exceeded, and how often did the violations occur. Human judgements can reduce the veracity of the data and these would include the customer deciding what day the transaction was completed and inputting that as the posted date. Additionally, the credit cardholders might incorrectly choose which merchant category code the transactions fall under. Although categorization has been provided by the credit card company, some transactions might not easily fall under the list provided.

Volume – the amount or scale of data

We will analyze only OSU data as their internal auditor for P-card transactions in 2014. This reduces data handling to 116, 031 rows (including the header), making analysis easy using Python Pandas module. All fields are relevant to our analysis. Agency Number and Name help focus on OSU, cardholder name and initial identify purchases and violations. Description, Vendor, and Merchant Category Code detect fraudulent transactions, while Amount, transaction date, and posted date show the details and extent of spending.

TASK 5

Summary of analysis results for 7 types of potential violations of P-card rules are shown below.

Internal Control	Result
1. User shall not spend more than $5,000 per transaction.	A total of 33 transactions exceeded the $5,000 per order limit.
2. User shall not spend more than $50,000 per year.	There were 127 cardholders who violated the year limit of $50, 000 during the year of 2014. The maximium yearly spending amount was $1,595,302, by cardholder G. Hines, followed by $1,182,094 of M. Tornakian.
3. User shall not spend more than $10,000 per month without approval.	Monthly limit was violated 457 times across the year. Notably, January maximum spending is $149, 700 by cardholder G. Hines.
4. An amount more than $5,000 should not be split between two or more swipes of the	A total of 66 transactions were splitted into two or more swipes but totaled more than $5, 000.

card by the same person.	
5. Count how often each individual purchaser did not follow the previous restriction. Sort this in descending order by the count.	With those individuals split purchasing by swiping a few times, G. Hines, M. Tornakian, and J. Kindschi have violated 18, 11, 8 times respectively.
6. Purchases should not be split between two or more cardholders.	Same-vendor purchases that are splitted between several cardholders with total amount over the $5,000 limit took place 41 times.
7. Transactions are prohibited for expenses for food and mileage while traveling. A per diem for food expenses and mileage may be claimed using a travel voucher.	A total of 90 food related transactions took place on those traveling days for some purchasers during the course of the year.

There are some notable types of prohibited purchases:

(1) Monthly/Yearly/Per Transaction Limits

We notice a few cardholders were spending way over the limit. Both G. Hines and M. Tornakian spent over $1 million per year on their P-cards, which are over 20 times of the $50, 000 year limit. The fact that monthly limit was violated 457 times was also a read flag. Even the per transaction limit was only exceeded 33 times, the total consumption violations are quite concerning.

(2) Split Purcahse

Split purchasing through swiping several times either by the same card or by different cardholders was not rare. In particular, G. Hines and M. Tornakian splitted more than 10 times using their own cards.

(3) Food Expenses

Restrictions on food expenses were violated 90 times by 31 cardholders.

Conclusion:

Even though potential internal control violations do not mean they are actual violations, there should be acceptable reasons and solid documentations associated. Above analysis can show a considerable chance of misconduct due to the frequency and high overspending amount. We suggest to undertake an immediate inspection to investigate potential fraudulent activities on higher-risk cardholders, and cancel or adjust their spending limit if needed.