

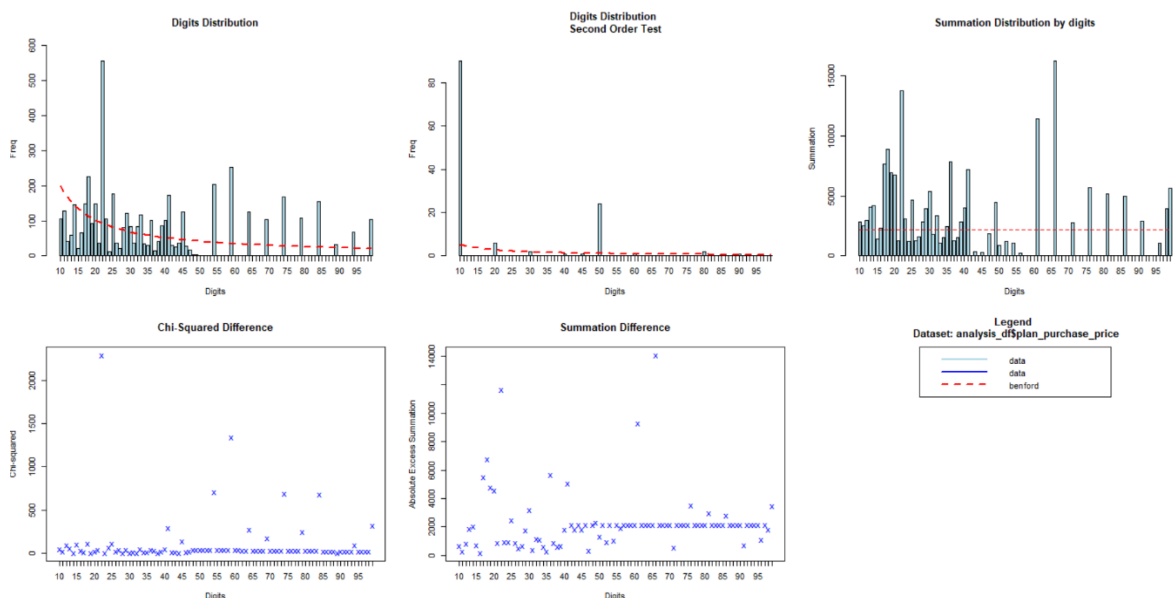
Question 1:

1. Please review the integrity of the data. Do you notice any data anomalies or idiosyncrasies? If so, please describe them.

Data unit testing was applied to the datasets. Most of the datasets had NULL values that were cleansed. For the merchant's data, there appeared to have been some additional pre-processing applied with a key based on each tore. Regex was used to remove the erroneous characters. Additionally, there appeared to be mismatches or data entry issues throughout the dataset that were cleansed before processing.

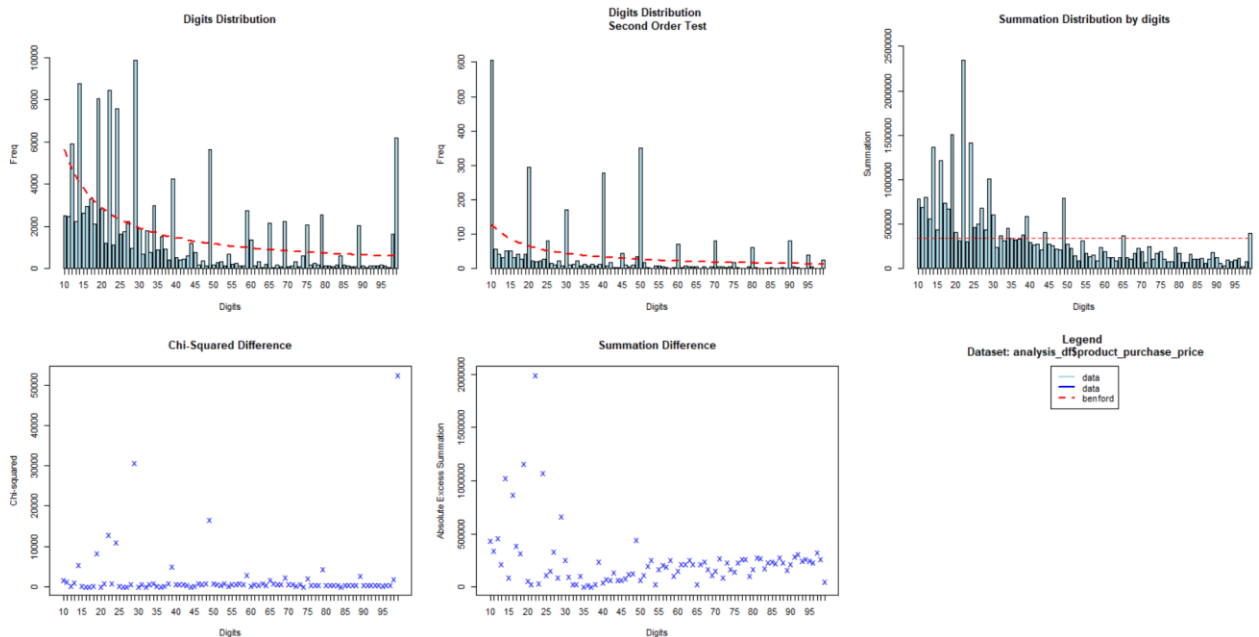
Additionally, forensic accounting techniques were applied to the data to review the integrity of data resources. The main form of analysis concerns Benford's Law. The law suggests that there is an underlying distribution in the frequencies of digits in numbers. If the frequencies of digits vary significantly from the expected distribution, then there may be a systematic bias in the dataset.

The following two sets of graphs were created from plan purchase price and product purchase price respectively.



The initial digit distribution suggests that the dataset does somewhat conform to Benford's Law. But, there is some worrisome distribution at the tail of the Digits Distribution. The Second Order test reaffirms this concern. The second order tests observe the difference of digits that are sorted and is a test that is weighted more than the digit distribution. Additionally, the Summation distribution is problematic. If Benford's Law held strongly, it would appear more uniform. However, there are spikes at regular intervals. These spikes suggest that there is sporadic pricing. One would expect some deviation in this distribution when looking at prices

though. It would be understandable if prices spiked with numbers including zeros. However, the variation is enormous and suggests there could be quality issues made in the largest prices in the dataset.



As for the product purchase price, the data appears to generally conform more nicely to Benford's Law. The Digits Distribution does have some variation, but it generally adheres. Likewise, the second order test follows as well. The variation in the distribution allows follows the expectation in pricing data. The peaks that deviate are associated with zeros that further suggests more confidence in the data. As for the Summation test, the distribution is far more uniform with some variation from the law with the smaller digits making up larger prices in the distribution. While this may be a concern here, this analysis suggests that it is not overly problematic for product purchase prices. However, there may be issues in plan purchase price. The data that deviated most significant has been highlight in the "analysis_df_benford.csv" file in the Data Unit Test directory.

2. Calculate attach rate overall and by merchant per month on a unit and dollar basis. Pick two merchants that are outliers (high attach rate and low attach rate) to do a SKU level analysis
 - A. Overall Attach Rate

General Attach Rate

3%

B. By Merchant By Month (Dollars)

Month	Merchant	Monthly Attachment Rate
Mar	CarStereolIsland	1.1172%
Mar	ElectricSkatePark	1.4031%
Mar	FitnessWareables	2.1167%
Mar	HomeSecurityMart	1.9369%
Mar	HVAC4U	1.7182%
Mar	Photos4Grandma	0.7075%
Mar	PowerMassagerPros	0.8247%
Mar	PowerWashersDirect	0.0144%
Mar	RefurbPCLand	1.5855%
Mar	SecurityCamDirect	3.6466%
Mar	SiiickHeadphones	2.4638%
Mar	SkateboardsUSA	1.2852%
Apr	CarStereolIsland	0.8842%

Apr	DietTrackers	0.2444%
Apr	ElectricSkatePark	1.6387%
Apr	FitnessWareables	2.0524%
Apr	HomeSecurityMart	1.6490%
Apr	HVAC4U	1.4212%
Apr	LuxuryWatchDepot	0.1053%
Apr	Photos4Grandma	0.6687%
Apr	PowerMassagerPros	1.0380%
Apr	PowerWashersDirect	0.2993%
Apr	RefurbPCLand	8.3660%
Apr	SecurityCamDirect	4.1915%
Apr	SiiickHeadphones	3.8998%
Apr	SkateboardsUSA	1.0207%
May	CarStereolIsland	0.8697%

C. Per Merchant Per Month (Units)

Month	Merchant Name	Monthly Attachment Rate By Unit
Mar	CarStereolIsland	16.07547%
Mar	ElectricSkatePark	9.33333%
Mar	FitnessWareables	10.54767%
Mar	HomeSecurityMart	6.36475%
Mar	HVAC4U	6.70860%
Mar	Photos4Grandma	4.14428%
Mar	SecurityCamDirect	42.27829%
Mar	SiiickHeadphones	8.36237%
Mar	SkateboardsUSA	2.41273%
Apr	ElectricSkatePark	9.37500%
Apr	FitnessWareables	9.62199%
Apr	HomeSecurityMart	5.22134%
Apr	LuxuryWatchDepot	1.04634%

Apr	Photos4Grandma	5.04946%
Apr	PowerMassagerPros	5.32351%
Apr	SecurityCamDirect	32.61738%
Apr	SiiickHeadphones	10.76417%
Apr	SkateboardsUSA	1.76056%
May	ElectricSkatePark	13.66460%
May	FitnessWareables	9.38101%
May	HomeSecurityMart	6.17549%
May	LuxuryWatchDepot	3.92157%
May	Photos4Grandma	5.22179%
May	SecurityCamDirect	11.39019%
May	SiiickHeadphones	8.90824%
May	SkateboardsUSA	1.80075%

D. SKU Analysis

Merchant	Plan	Plan Purchase Price Sum
RefurbPCLand	10001-laptop-base-replace-1y	1569.24
RefurbPCLand	10001-laptop-base-replace-2y	2825.20
PowerWashersDirect	10001-misc-elec-base-replace-1y	655.88
PowerWashersDirect	10001-misc-elec-base-replace-2y	1486.82
PowerWashersDirect	10001-misc-elec-base-replace-3y	876.94
RefurbPCLand	10001-misc-elec-base-replace-3y	244.95

Merchant	Plan	Plan Purchase Price Mean
RefurbPCLand	10001-laptop-base-replace-1y	20.64789
RefurbPCLand	10001-laptop-base-replace-2y	35.31500
PowerWashersDirect	10001-misc-elec-base-replace-1y	20.49625
PowerWashersDirect	10001-misc-elec-base-replace-2y	61.95083
PowerWashersDirect	10001-misc-elec-base-replace-3y	97.43778
RefurbPCLand	10001-misc-elec-base-replace-3y	48.99000

In terms of plans being purchased through the RefurbPCLand, the laptop warranties seem to be popular. It seems like an intermediate length warranty is preferred. My intuition suggests two things. Firstly, if laptops are refurbished, people may be more concerned and are willing to spend the extra money for the warranty. Additionally, it corresponds to Moore's Law and people may be seeking a new computer around that time. That may explain why people would not want a longer term one.

For PowerWashersDirect, there seems to be less spend. The same trend of the 2-year warranty being most purchased holds. But the price point for the warranty is far higher. It may be that the warranties for these products are being priced out and are thus not being consumed.

6. Please share any feedback you have on the exercise and how we can make it better!

It would be useful to have terms defined and to be more specific in language. For instance, a 3-page PowerPoint did not have a clear meaning. I interpreted to be 3 slides per page.