

**Project Report**

**Course ID: CPS 3320\*01**

**Project 1: Project Documentation**

**Student: Rainer Luis**

**Instructor: Robert Domanski**

**Date: 03/01/2023**

**Background:** As a data analyst at my work, I have encountered many situations where a faster and easier to use processing tool is needed. Usually, I would do a lot of inventory performance analysis where two different inventory reports from two different months usually consecutive are analyze against each other to find the efficiency of counts as well as the accuracy of our transaction process. With the use of python my objective is to perform this same analysis that I do at work with excel but this time creating a python script that will calculate the cost as well as qty differences from moth to month as well as create a pie chart with a percentage breakdown of the most volatile product.

**Hypothesis:**  My Hypothesis is that from week to week the higher difference in consumption will come from packaging material, there are many factors that may contributed to this, some of which are the lack of a waste accountability, the distribution of our packaging to different vendors, as well as the difference in size of the packaging, in order to be more precise in the analysis I would based in the asset cost to avoid qty difference errors.

**Data:** the data represents two inventory physical counts from my job, this physical count includes 2 warehouses, as well as 4 external vendors inventories. The source of the data are two physical counts perform using a Microsoft Access application that allow users to scan product into the system, contrast this product against live inventory and record this product into a physical count table, this table is then translated into a csv file for further analysis. The outside vendors have different process of counting but send their physicals which are further cleaned up and add to this csv dataset. The possible biases of this dataset would be drastic change in qty because of human error while counting, inconsistent items numbers been a combination of numbers and letters, as well as a complicated column and row structure with locations been added in the rows instead of the columns. A challenge in this dataset is that there’s a lot of items that have comas in their name and it confuses python reader. I had to clean up the data a lot before processing and decided to use cost as my based of calculation instead of qty to avoid excessive qty reporting based on different item sizes.

**Analysis/Methodology:**

1. The first step would be to import the dataset into excel, I would proceed to clean up some of the data to make sure my python reader can translate the data correctly. To be more specific I had to delete an entire column of my dataset which was item descriptions because of its excessive used of comas which confused the python reader.
2. I imported the csv file into python. This file consists of an inventory site column, item number, item type, asset value, inventory qty(previous month) physical qty (current month).
3. I used a dropna command to make sure my dataset was clear of any “na” values. And start calculating the inventory asset and physical asset values.
4. Two new columns were added. To a new dataset, the first of this column been the inventory asset values which was calculated by inventory qty \* the asset value of the item. The second column been the physical asset value was calculated by taking physical qty \* asset value.
5. Additional columns were then added, one of them been the variance qty witch would store the difference quantities of each item number per site by doing physical qty – inventory qty. the second column called difference consist of the difference in money between both inventories, this gets calculated by doing physical asset – inventory asset.
6. The newly created data frame then gets store into a new csv file named AnlayzedInventoryReport.csv.
7. To do the report a bit more visual created some variables for values In column item type and calculated the total value of it, to determine with item type was reporting the highest drastic change and contrast my hypothesis.

**Conclusion:** After all totals of item types were extracted a pie chart was then created to show the high changing item type

Chart, pie chart

Description automatically generated

The end pie chart shows that packaging indeed is the most volatile item, but WIP was close behind it, given me an inside that process in these two items must be reworked and enchanted. This result would support my hypothesis, but depending on the type of month percentages may drastically change, but packaging would always be the higher discrepancy issuer. The hypothesis most be further analyze in multiple instances to see if results are consistent and keep analyzing If WIP discrepancies get closer to 0.