**Business Intelligence Report**

**Pixysystems**

**Table of Contents**

[1. Introduction 2](#_Toc42272673)

[Overview 2](#_Toc42272674)

[Data Treatment: 3](#_Toc42272675)

[2. Discussion and Analytics 3](#_Toc42272676)

[Other visualizations: 14](#_Toc42272677)

[3. Conclusion 15](#_Toc42272678)

[4. References 16](#_Toc42272679)

[5. Appendix 17](#_Toc42272680)

# Introduction

## Overview

The purpose of the report is to draw useful and meaningful insights from data collected for the benefit of the organisation ‘Pixysystems’. The dataset is provided by the company, and involves data on the financial and operational management of the business. The concrete results drawn from this report will help the Chief Financial Officer of the company to make effective decisions which would improve the business both financially and operationally.

Data recently surpassed oil to be the most valuable resource in the world. With such growth in volume of data storage and transfer on a daily basis, it is becoming more and more important to analyse such data. This is done using big data analytics tools and techniques. Big data analytics entails analysing large amounts of data and identifying patterns and correlations to provide useful results to businesses [1]. In this project, large amounts of data is analysed to provide useful insights about the financial and operational aspect of the business.

The data is provided by the company ‘Pixysystems’. Pixysystems is a large manufacturer and distributer of toys all around the world. The data provided entails information about its employees, payroll, finished goods, inventory, costs, sales, and so on. Each of the heads mentioned above has a separate excel sheet, with different types of data. This data is then analysed on Python to assist the business in correcting inefficiencies and becoming more profitable.

## Data Treatment:

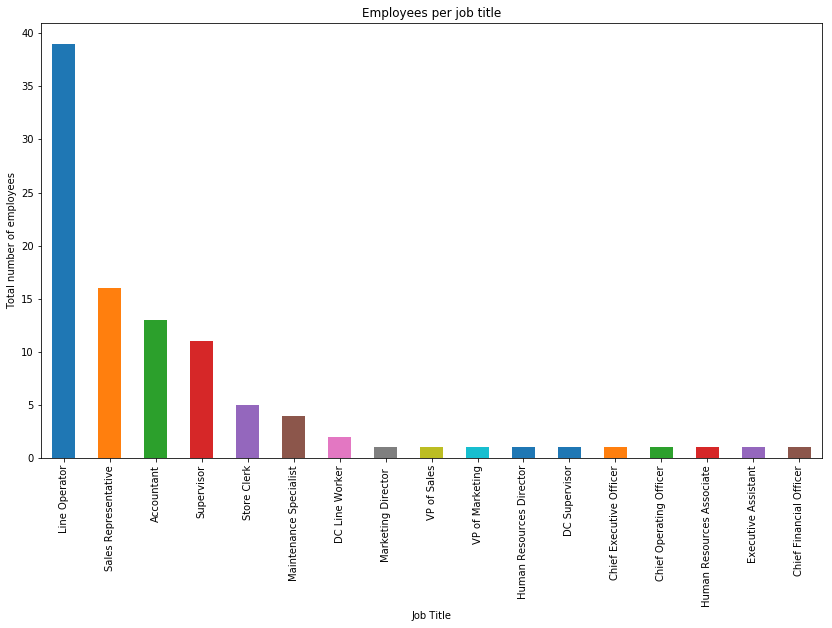
Various packages and techniques in python are used to analyse and visualize the data. Numpy and Pandas are used for analysis and reporting, whereas Matplotlib.pyplot and Seaborn as used for visualizations.

# Discussion and Analytics

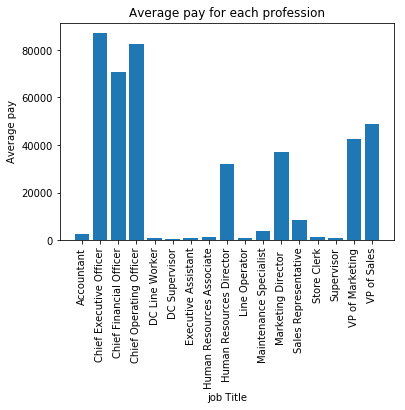
The dataset is analysed and visualized using Python to provide meaningful results and insights to the business. This analysis is done using Python. Firstly, the necessary packages are imported for the analysis and visualizations of the dataset. The necessary python packages needed for this project are Pandas, Matplotlib.pyplot, seaborn and Numpy. Once that is finished, then the dataset is loaded using pandas. After loading the dataset, the analysis is done in the following manner:

Since the dataset has multiple excel sheet, we first focus our attention towards the employees of the business and load the sheet which gives us information on the same. The first sheet loaded gives information on each employee of the company under the following heads: medical plan, savings plan, total annual pay, tax, savings, year-end bonus, and job title. This dataset has 22 columns and 100 rows.

The analysis first focuses on the total number of employees under each job title, which can be shown in the following manner:

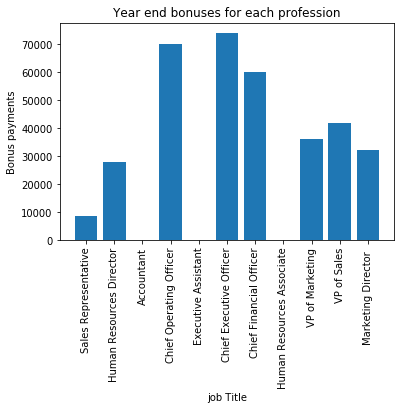


It is evident that Line operator has the highest number of employees, followed by sales representative and accountant. The analysis shows that the business hires about 39 line operators, 16 sales representatives, and 14 accountants. Next, the authors look at the average pay in each profession, which can be seen as follows:

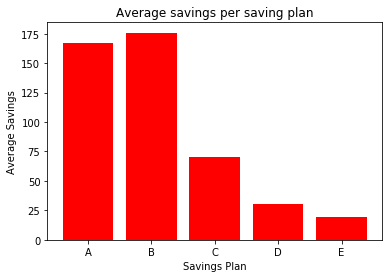


It is evident that there is a huge pay gap in the organisation. Chief executive officer, chief financial officer and chief operating officer earn way more than all of the company’s employees. DC supervisor earns the lowest. Thus, there is a huge pay gap between the highest and the lowest earners.

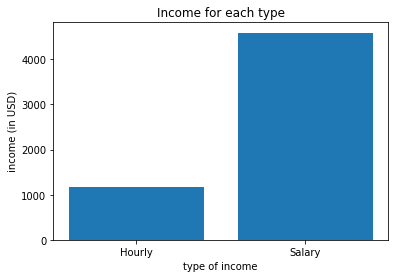
To now focus on differences in bonus payments, we look at the following graph:



Accountant, executive assistant and HR associate seem to enjoy zero bonus benefits in the organisation. It is evident that almost all employees receive huge end of the year bonuses, except accountants, line operators, executive assistant and human resource associates. Additionally, looking at average level of saving in each plan, we get the following result:



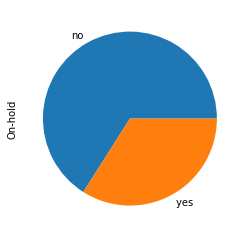
It is evident that plans B and A lead to the highest levels of average savings, whereas plan E leads to the lowest. it is evident that plans A and B, on average, tend to attain the highest amount saved, whereas plan E has the lowest.



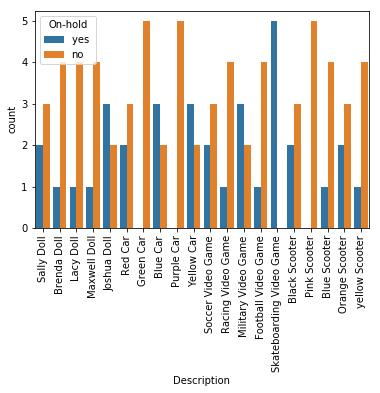
The above graph shows the vast difference in earnings among individuals who earn hourly vs individuals who earn a fixed salary. Individuals who earn a fixed salary earn a significantly larger amount than those who earn hourly, on average

The analysis now shifts to manufacturing of good. The dataset used provides information under the heads of Location of the product, type of product, quantity of product available at each location, whether the product is on hold or not, and the date the product was received from the factory. The dataset is loaded and exploratory analysis is run on the same.

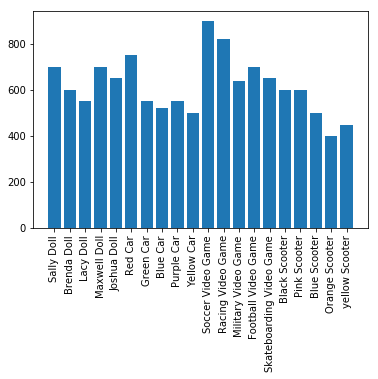
The analysis first focuses on how many products are on hold, which can be shown in the following manner:



The above visualization shows that about 34% of the products are on hold, while 66% are not on hold. Out of 100 products available, 34 of them are on hold. To determine which ones are on hold, we focus on the following visualization:

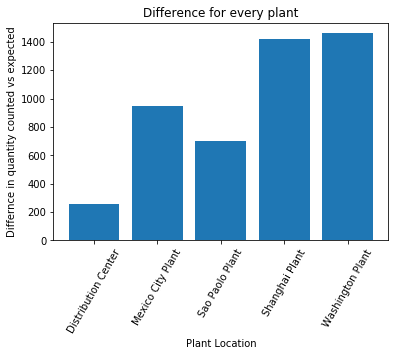


The above graph shows how much quantity of each item is on hold out of the total quantity of each item. It is evident that many products like skateboarding video games, yellow car, blue car, etc have more units on hold than those not on hold. The analysis now looks at total quantity available for sale per item:

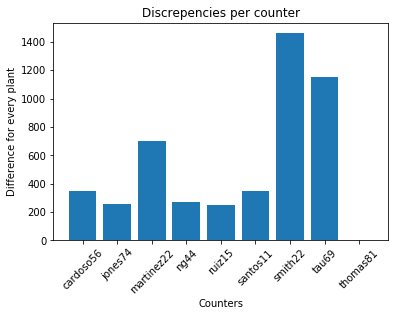


It is evident that soccer video games are the most abundant items in the inventory, followed by racing video games and military videogames, whereas orange and yellow scooter are the least abundant.

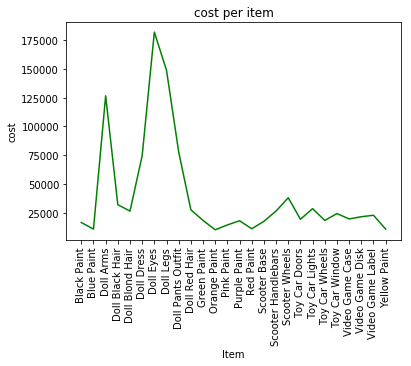
The analysis now shifts from finished goods to inventory analysis. The dataset is first imported which gives information on the expected vs actual inventory count in each location for each item. After doing exploratory analysis, we find that the dataset has 10 columns and 204 rows. The total quantity available with the firm is 205175 units. To look at discrepancies between the actual quantity vs expected quantity, we look at the difference of each location, as follows:



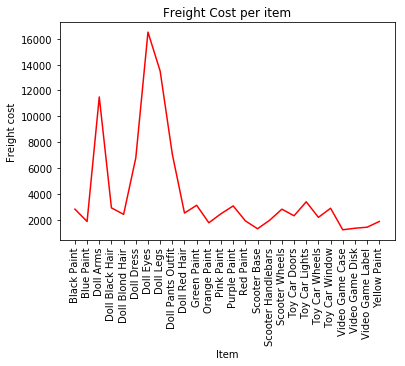
To analyse the discrepancies and inefficiencies in each plant, the difference between actual vs expected quantity is looked at. Washington Plant tends to have the highest level of inefficiency, whereas Distribution centre has the lowest. It is evident that discrepancies are highest in the Washington and Shanghai plants, whereas it is the lowest in the distribution centre. Thus, this means that there is some for of inefficiency which exists in the Washington plan in terms of either counter or delivery. To check for whether these discrepancies occur due to counting, we look at the following graph:



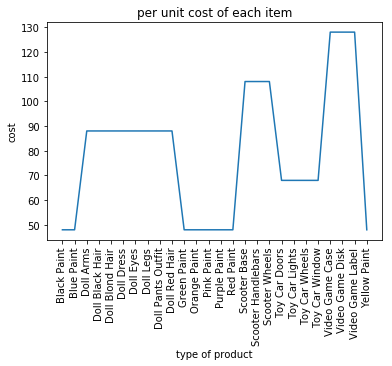
To look for why such discrepancies exist, we look for the accuracy of each counter, and find that smith 22 shows the highest level of inefficiency. It is evident that smith 22 has the highest difference value between actual quantity and expected quantity, followed by tau 69. Next, we focus on the cost structure of the firm. The data which involves the cost of each item is uploaded using pandas, and can be shown using the following line chart:



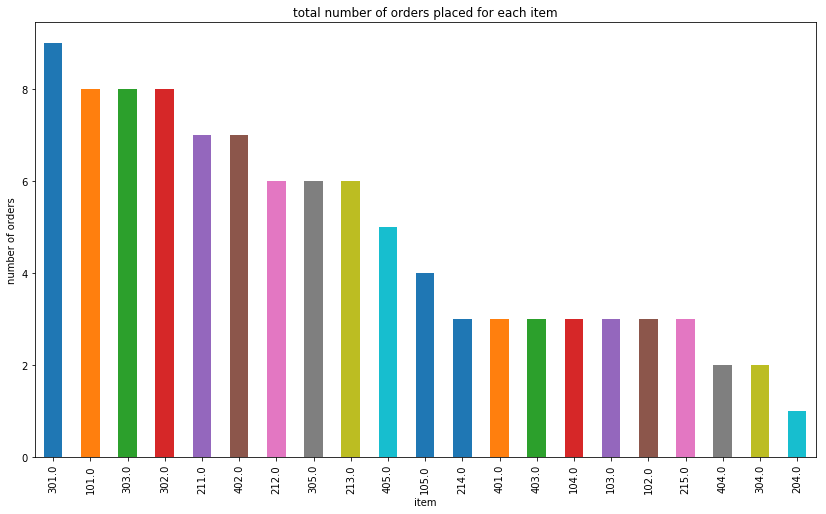
The above graph shows that products like doll arms, doll eyes, doll dress, and doll legs cost the highest among all items. While, paint products cost the lowest. Thus, manufacturing a doll seems to be highly expensive.



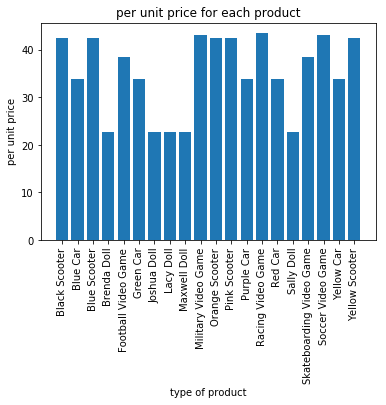
The graph above shows the freight cost involved with each product. It is evident that doll related products have a very high freight cost as well. To take a closer look, per unit cost is taken into consideration as follows:



It is evident that video games cost the highest per unit, while paint costs the lowest. Now we focus on the sales and price of the goods in the following manner:



The above graph shows the number of orders placed for each item. It is evident that item number 301 has the highest demand.



The above graph gives the per unit price of each finished product. This, when compared with the per unit cost of each item, we see that dolls are the least profitable products, whereas scooters and video games are highly profitable. This, when compared with per unit cost, we find that dolls are least profitable.

## Other visualizations:

Some visualizations or analytics were not added in the report or done by the author. One such visualization is the total cost vs total revenue. Another visualization like this would be profit vs revenue. These visualizations can only be created if information about how the final product is manufactured and the cost and revenue of each is given.

Similarly, Financial analysis like current ratio, quick ratio, debt to equity ratio can also be done to provide a better holistic approach of the financial position of the company. These analyses can only be conducted if financial information and balance sheets of the company are provided.

Moreover, if the data about the total and average product of the company was given, it would help in assessing the efficiency of the company.

# Conclusion

Benefits of using data analytics are tremendous, and more and more organisations are adopting such techniques and technologies [2]. The meaningful insights drawn from the above report can be used to correct the inefficiencies which exist with the firm and thus make it more profitable. These results can help the business make corrections, both in the business operations as well as in the pay or financial structure. The author provides some recommendations, which are as follows:

1. The pay gap in the company is incredibly high. Moreover, many employees do not enjoy year end bonuses, while some receive bonuses extravagantly. This can reduce the morale in the company. Thus, employees need to be given bonuses fairly.
2. Saving plans need to be revalued since only plans A and B tend to be highly effective.
3. 34% of the total goods are put on hold. An investigation needs to be conducted to find the reason behind it.
4. Washington plant has the highest discrepancy and inefficiency. An investigation needs to be established to correct this discrepancy. Moreover, Smith 22 seems to be highly inefficient, which needs to be corrected.
5. Dolls are the least profitable among all products. Thus, CFO must find a way to increase price or reduce costs of dolls to continue its production.

The above recommendations can assist the business in correcting inefficiencies and become more profitable.

# References

|  |  |
| --- | --- |
| [1] | J. R. Alam, A. Sajid and R. Talib, “A Review on the Role of Big Data in Business,” *International Journal of Computer Science and Mobile Computing,* 2014. |
| [2] | I. Ajah and H. Nweke, “Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications,” *MDPI,* 2019. |

# Appendix

**Python Code:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

#importing dataset1

xls = pd.ExcelFile(r'C:\Users\VIDHI\Desktop\fictionn.xlsx')

df1 = pd.read\_excel(xls, 'S1')

df1.info()

df1.head()

df1.describe()

#%%

job = df1['Job Title']

print(job)

#%%

#frequency of type of job

jobcount = job.value\_counts().plot(kind='bar', figsize = (14,8))

plt.xlabel('Job Title')

plt.ylabel('Total number of employees')

plt.title('Employees per job title')

#%%

plt.bar(job, df1['Total Pay'],)

plt.xlim(['Line Operator', 'VP of Sales'])

plt.show()

#%%

#grouping according to job title

avg = df1.groupby('Job Title').mean()

#%%

#average pay according to profession

plt.bar(avg.index, avg['Total Pay'])

plt.xticks(rotation = 90)

plt.xlabel('job Title')

plt.ylabel('Average pay')

plt.title('Average pay for each profession')

#%%

#bonus for every position

plt.bar(df1['Job Title'], df1['Year End Bonus'])

plt.xticks(rotation = 90)

plt.xlabel('job Title')

plt.ylabel('Bonus payments')

plt.title('Year end bonuses for each profession')

#%%

#average bonus accoridng to profession

plt.bar(avg.index, avg['Year End Bonus'])

plt.xticks(rotation = 90)

#%%

#grouping based on savings plan

sv = df1.groupby('Savings Plan').mean()

print(sv)

#%%

#level of average savings in every plan

plt.bar(sv.index, sv['Savings'], color='red')

plt.xlabel('Savings Plan')

plt.ylabel('Average Savings')

plt.title('Average savings per saving plan')

#%%

#level of average pay per savings plan

plt.bar(sv.index, sv['Total Pay'])

#%%

#importing dataset2

df2 = pd.read\_excel(xls, 'Finished Goods On-Hand')

#%%

#frequency of on hold

onhold = df2['On-hold'].value\_counts()

x = onhold.plot(kind='pie')

#%%

#on hold analysis

yyx = sns.countplot(x='Description', hue = 'On-hold', data = df2)

yyx.set\_xticklabels(yyx.get\_xticklabels(), rotation=90)

#%%

#total quantity per item

plt.bar(df2['Description'], df2['Quantity'])

plt.xticks(rotation = 90)

#%%

#grouping based on average

averq = df2.groupby('Description').mean()

print(averq)

#%%

#average quantity per item

plt.bar(averq.index, averq['Quantity'])

plt.xticks(rotation = 90)

#%%

#importing data3

df3 = pd.read\_excel(xls, 'Inventory Count')

#%%

#total quantity

m = df3['Quantity Counted'].sum()

print(m)

#%%

#grouping based on location

n = df3.groupby('Detailed Location').sum()

#%%

#difference

plt.bar(n.index, n['Difference'])

plt.xticks(rotation=60)

plt.xlabel('Plant Location')

plt.title('Difference for every plant')

plt.ylabel('Differnce in quantity counted vs expected')

#%%

b = df3.groupby('Counter').sum()

#%%

#counter

plt.bar(b.index, b['Difference'])

plt.xticks(rotation = 45)

plt.xlabel('Counters')

plt.ylabel('Difference for every plant')

plt.title('Discrepencies per counter')

#%%

#importing dataset 4

df4 = pd.read\_excel(xls, 'Quantt')

#%%

df4.head()

#%%

#total cost

cost = df4['Total Cost'].sum()

print(cost)

#%%

#grouped based on description

total = df4.groupby('Description').sum()

print(total)

#%%

#cost of each material

plt.plot(total.index, total['Total Cost'], color='green')

plt.xticks(rotation = 90)

plt.xlabel('Item')

plt.ylabel('cost')

plt.title('cost per item')

plt.show()

#%%

#freight cost

plt.plot(total.index, total['Freight Cost'], color='red')

plt.xticks(rotation = 90)

plt.xlabel('Item')

plt.ylabel('Freight cost')

plt.title('Freight Cost per item')

plt.show()

#%%

#per unit cost

plt.plot(total.index, total['per unit'])

plt.xticks(rotation = 90)

plt.xlabel('type of product')

plt.ylabel('cost')

plt.title('per unit cost of each item')

#%%

#importinf data5

df5 = pd.read\_excel(xls, 'Sales')

#%%

products = df5['Itemm'].value\_counts()

print(products)

#%%

#total orders of each product

products.plot(kind='bar', figsize=(14,8))

plt.xlabel('item')

plt.ylabel('number of orders')

plt.title('total number of orders placed for each item')

#%%

#importing dataset for payroll

df6 = pd.read\_excel(xls, 'Payroll')

#%%

#grouping based on employee type

gld = df6.groupby('Employee Type').mean()

print(gld)

#%%

#average pay per employee type

plt.plot(gld.index, gld['Pay'])

plt.xlabel('type of income')

plt.ylabel('income (in USD)')

plt.title('Income for each type')

#%%

#importing dataset 7

df7 = pd.read\_excel(xls, 'per')

df7.head()

#%%

xxx = df7.groupby('Finished Items').sum()

print(xxx)

#%%

#per unit price

plt.bar(xxx.index, xxx['Suggested Distributor Price'])

plt.xticks(rotation = 90)

plt.xlabel('type of product')

plt.ylabel('per unit price')

plt.title('per unit price for each product')