EBUS 3030

Assignment 2  
Group: POE

Team members:

Sander Fabian Visvaseelan (c3418982)  
Oh Chen Neen (c3415898)  
Victor Chua Jia Zhi (c3418248)  
Brent Lee Ting Zhen (c3415641)  
Bui Ho Minh Tho (c3415439)  
Wei Yang Cheong (c3415898)

Table of Contents

[1.0 Introduction 2](#_Toc151726515)

[1.1 Problem Statement 2](#_Toc151726516)

[1.2 Proposed Solution 2](#_Toc151726517)

[2.0 Deriving an Answer 2](#_Toc151726518)

[2.1 Assumptions 2](#_Toc151726519)

[2.2 Identifying the Anomalies 2](#_Toc151726520)

[2.3 Cleaning the Database 3](#_Toc151726521)

[2.4 Creating a Datamart for Analysis 3](#_Toc151726522)

[2.5 Creating a Dashboard for Visualization 3](#_Toc151726523)

[2.6 Findings 3](#_Toc151726524)

[2.6.1 Staff KPIs 3](#_Toc151726525)

[2.6.2 Item KPIs 3](#_Toc151726526)

[2.6.3 Incentive Changes 3](#_Toc151726527)

[2.6.4 Branch closure 3](#_Toc151726528)

[3.0 Conclusion 4](#_Toc151726529)

[3.1 Suggestions 4](#_Toc151726530)

[3.2 Benefits 4](#_Toc151726531)

[3.3 Closing statement 4](#_Toc151726532)

[4.0 Appendices 5](#_Toc151726533)

[4.1 Datamart Model 5](#_Toc151726534)

[4.1.1 Datamart Diagram 5](#_Toc151726535)

[4.1.2 DataMart Definitions 5](#_Toc151726536)

[4.1.3 DataMart Implementation 5](#_Toc151726537)

[4.2 Anomaly Detection 5](#_Toc151726538)

[4.2.1 Solutions for cleaning the Data 5](#_Toc151726539)

[4.3 ETL Processes 5](#_Toc151726540)

[4.4 Assumptions On staffID 5](#_Toc151726541)

[4.5 Implementation of a Dashboard 5](#_Toc151726542)

[4.6 Base Analysis 5](#_Toc151726543)

[4.6.1 Sales staff KPI 5](#_Toc151726544)

[4.6.2 Item KPIs 5](#_Toc151726545)

[4.6.3 Branch KPIs 5](#_Toc151726546)

[4.6.4 Incentive Measures 5](#_Toc151726547)

# Introduction

Hello Jessica Smotherington,

We wish to present this business intelligence report to help you with your enquiries. With the data provided, we have cleaned and processed it to a Datamart that would enable the analysis of the areas you have expressed interest in. This report aims to answer your questions and suggest possible directions for “Bits&Bobs” to take going forward.

## 1.1 Problem Statement

“Bits&Bobs” is currently struggling to identify the correct measures to take to reduce costs and help the business thrive in the long term. Jessica Smotherington has also made suggestion for an alternative customer sale incentive that could potentially offset the costs associated with the previous discount system. “Bits&Bobs” is also operating across 10 branches and seek to downsize to save operational costs. Jessica Smotherington also wants to identify the top salesperson based on various KPIs and the marketability of the items in their inventory.

## 1.2 Proposed Solution

We are here to solve Jessica’s current struggles with the implementation of a dashboard based on the DataMart created with the data provided for 2022. We aim to explore the new customer point systems in conjunction with the prior 5% discount system. An analysis of the items, staff and the branches, was conducted to steer “Bits&Bobs” through this economic downturn and ensure the longevity of the business with facts and data.

# 2.0 Deriving an Answer

## 2.1 Assumptions

Based on the database presented, a few assumptions about the data were made:

1. Every staff ID is unique to each staff. (Refer to Appendix 4.4)
2. All fields should have values.
3. Receipts should be associated to 1 customer.
4. Customer ID should be unique.
5. If a customer has purchased more than 5 items, a 5 % discount will be discounted in the row total and placed under the discounted row total.
6. Receipts should not span over multiple dates.
7. Each receipt’s transaction row can only be associated to a singular staff.

## 2.2 Identifying the Anomalies

The database was sent through a variety of checks to identify any anomalies in the data. This process is illustrated within Appendix 4.2.

## 2.3 Cleaning the Database

With the anomalies identified, the database was cleaned using SSIS. This process is illustrated within Appendix 4.3.

## 2.4 Creating a Datamart for Analysis

A Datamart was designed and created for use in analysis, the details are highlighted in Appendix 4.1. The implementation of the Datamart can be observed in 4.1.4.

## 2.5 Creating a Dashboard for Visualization

The implementation of the dashboard can be observed in 4.5.

## 2.6 Findings

Through the dashboard the many areas of interest were addressed. To summarize the findings of the base analysis:

### 2.6.1 Staff KPIs

**Top salesperson overall (Revenue):   
Staff: Savannah Jones  
Branch: Wollongong**

**Top salesperson overall (Qty):   
Staff: Samuel Anderson  
Branch: Lismore**

**Top salesperson overall (Revenue):   
Staff: Samuel Anderson  
Branch: Lismore**

This is explored further in appendix 4.6.1

### 2.6.2 Item KPIs

**Best:**

1. **Cordless Drill kit**
2. **Socket Set**
3. **Grinder**

**Worst:**

1. **Drill bit 2mm**
2. **Drill bit 3mm**
3. **Drill bit 4mm**

**This is explored further in appendix 4.6.2**

### 2.6.3 Incentive Changes

**From 2022s data, it would be beneficial to switch to the new point system incentive as there would be a cost reduction of ~$421,000. This however is merely arithmetic and might not necessarily translate as it does not consider consumer sentiment to a change like this.**

### **2.6.4 Branch closure**

The data indicates that Wagga Wagga is the top performing branch while Wollongong performs the worst. On closer inspection, the top performing branches also have higher staff associated with them and as such a directly proportional pattern can be seen. Based on staff to profit metrics, Broken hill is the least efficient branch. This could be due to various circumstances, from staff morale to a generally lower demand of hardware tools in Broken Hill relative to the other branches.

Unfortunately, projections could not be performed reliably due to a lack of enough historical data. With the information provided, the data would point to the closure of the Wollongong branch solely based on profits alone, however, the best salesperson overall, Savannah Jones may seek other employment due to this closure.

This analysis can be explored further in appendix 4.6.3

# 3.0 Conclusion

## 3.1 Suggestions

**Data Suggestion:**  
In the recommendation, we would like to propose a solution to Bits&Bobs to record the staff middle name and the staff email to prevent data issues and confusion on the analysis team. We would like to suggest that “Bits&Bobs” stores more historical data so that analysis can be conducted on a larger snapshot yielding results with greater precision.  
**Incentives for Sales Staff:**We recommend continuing the structured incentive program to motivate the sales team. This program should include the three KPIs highlighted.

**Underperforming Items:**Considering the 3 lowest sold items are the Drill bits ranging 2mm – 4mm, there maybe incentive to sell them as a pack.

**Customer Sales Initiative:**The point system will ultimately save costs for the organization. However, customer resistance may prove to reduce the effective savings. Surveys could be sent out to

**Branch Suggestions:**Consider trying to flatten the sales personal across and see if the staff disparity is causing the rift in the revenues earned by Wagga Wagga and Wollongong by distributing manpower budgets to lower staffed branches. If the need to downsize is dire, close the lower performing branches like Wollongong and relocate the sales personal wherever possible.

## 3.2 Benefits

By tackling the Areas mentioned above, “Bits&Bobs” should be able to transition into a better system to incentivise customer purchase and continue to encourage and bolster sales staff morale. By reducing costs and pivoting certain products into bundles, profits should increase by the reduction of operational costs and increase in revenue generated.

## 3.3 Closing statement

In conclusion, the efforts to keep the business running smoothly in the long term is noticed. This report has hopefully given “Bits&Bobs” a chance at a more informed decision. The commitment to keeping sales staff morale high is integral for the longevity of the business considering they are the foundation to the business’s success.

# 4.0 Appendices

## 4.1 Datamart Model

### 4.1.1 Datamart Diagram

A diagram of a company

Description automatically generated

### 4.1.2 DataMart Definitions

**Fact table (Sales)**

The fact table (sales) is designed to include the most appropriate attributes to derive meaningful insights into Bits & Bob's overall performance to answer Jessica Smotherington questions. The fact table (sales) consists of:

Primary and Foreign keys:

* salesId: the salesId serves as a surrogate primary key and is a system-generated identifier with no business meaning like the natural keys. This would allow data analysts to identify each row with ease. (Surrogate primary key)
* customerId: Customer who is associated with the sale transaction row. (Foreign key relationship with customer dimension table)
* date: The date of purchase in the sales transaction row. (Foreign key relationship with Time dimension table)
* officeId: The location ID of where the sales transaction row took place. (Foreign key relationship with Office dimension table)
* staffId: The staff who handled the sales transaction in the row. (Foreign key relationship with Staff dimension table)
* itemId: The item ID sold in each sales transaction row. (Foreign key relationship with Item dimension table)

Metrics:

* receiptId: We have decided to place receiptId in the metric area and not a dimension table since there is no additional information associated with the receiptId, unlike the previous data mart.
* Transaction row: The transaction row metric would show the multiple rows associated with each receipt ID.
* Item price: The item price would show the price tag associated with the Item ID.
* Item quantity: The quantity would show how many items are purchased in each transaction row.
* Row total: The row total would show the total cost of each transaction row.
* Discounted row total: The discounted row total would show whether there is a 5% discount on the row total when five or more unique items are purchased in each receipt and if not, it would just place the row total amount.

**Time dimension table**

The time dimension table contains a compilation of information regarding the date, which enables the data analysts to review additional information about the specific date based on the primary key of the table.

The time dimension table consists of:

* date – primary key
* day
* month
* year

**Office dimension table**

The office dimension table contains a compilation of information regarding the different locations in the organisation, which enables the data analysts to review additional information about the office based on the primary key of the table.

The office dimension table consists of:

* officeId – primary key
* location

**Customer dimension table**

The customer dimension tables contain a compilation of information regarding the customer, which enables the data analysts to review additional information about the particular customer based on the primary key of the table.

The customer dimension table consists of:

* customerId – Primary key
* first name of the customer
* surname of the customer

**Staff dimension table**

The staff dimension table contains a compilation of information regarding the staff in the organisation, which enables the data analysts to review additional information about the particular staff based on the primary key of the table.

The staff dimension table consists of:

* staffId – primary key
* first name of the staff
* surname of the staff

**Item dimension table**

The item dimension table contains a compilation of information regarding the variety of items sold by the organisation, which enables the data analysts to review additional information about the item based on the primary key of the table.

The item dimension table consists of:

* itemId – primary key
* description

The DataMart schema was developed based on the assumptions that the data processed was cleaned of any data anomalies. The star schema was chosen due to the simple nature of the table, and we felt it was best suited for this analysis.

### 4.1.3 DataMart Implementation

After the data is cleaned, the implementation of the data mart is created and populated.

**Time dimension table:**

A screenshot of a computer

Description automatically generated

**Customer dimension table:**

A screenshot of a computer

Description automatically generated

**Office dimension table:**

A computer code with text

Description automatically generated

**Staff dimension table:**

A computer code with text

Description automatically generated with medium confidence

**Item dimension table:**  
A computer code with text

Description automatically generated with medium confidence

**Fact table (Sales):**  
A computer screen shot of a computer code

Description automatically generated

A screenshot of a computer program

Description automatically generated

**SQL-generated database diagram**

A screenshot of a computer

Description automatically generated

## 4.2 Anomaly Detection

**Clean: Check whether there a null or empty values in the customer first name or surname.**

A screenshot of a computer code

Description automatically generated

**Clean: Check whether all unique items have multiple item descriptions.**

A screen shot of a computer code

Description automatically generated

**Clean: Check whether the staff ID contains more than one staff first name and surname**

A screenshot of a computer code

Description automatically generated

**Clean: Check whether the customer ID contains more than one customer first name and surname**A computer screen shot of a code

Description automatically generated

**Clean: Check whether the item price multiple by quantity is equivalent to the row total.**A close-up of a computer screen

Description automatically generated

**Clean: Check whether the office id shares a different office location.**A computer code with text

Description automatically generated

**The first anomaly detected (Row: 109678): Row 109678 has a null value in the row total.**

A screenshot of a computer

Description automatically generated

**Query:**

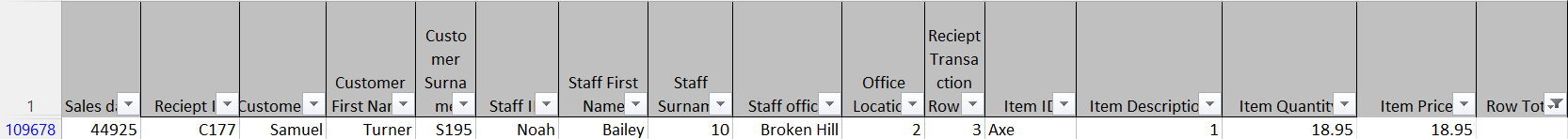
A close-up of a text

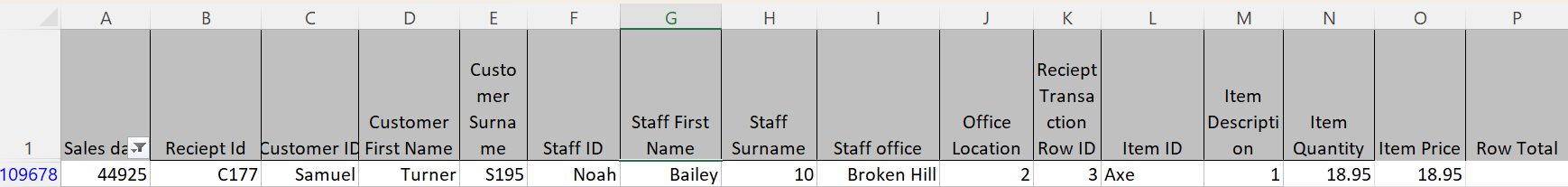
Description automatically generated

**Results:**

A close up of a white grid

Description automatically generated

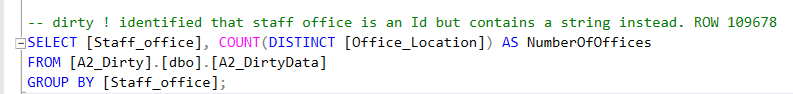


**The second detected (row: 109678): Under the sales date column, there is an invalid date format detected**A screenshot of a computer

Description automatically generated

**Third anomaly detected (row: 109678): Under the staff office column, it is supposed to be an office ID but instead it is a string named “Broken Hill” and under the office location, it is supposed to have the name of the office location, not the number “2”. Furthermore, “Broken Hill” seems to belong to the staff office ID “10.”**

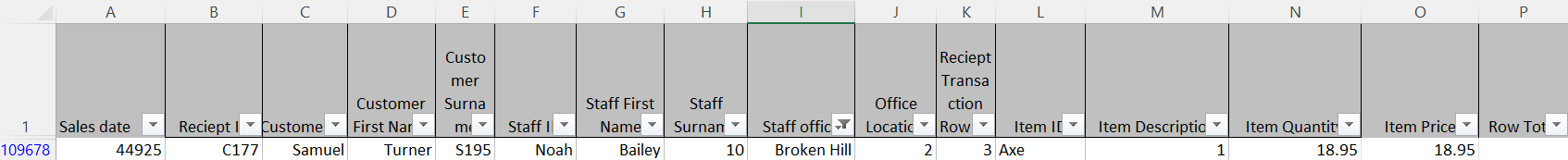
**Query:**



**Results:**

A screenshot of a computer

Description automatically generated



**Fourth anomaly detected: (Row: 109678):**

**Under the transaction row, it starts with 3 instead of 1.**

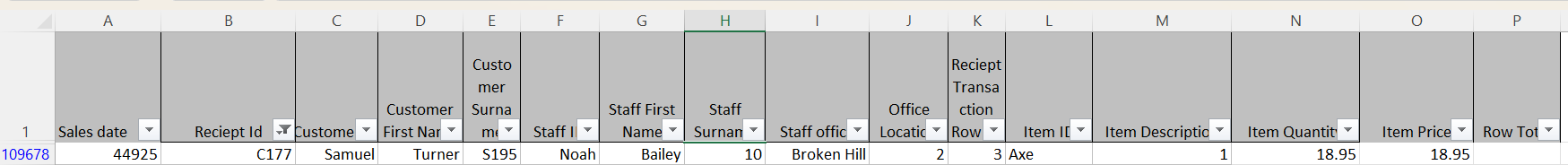
**Query:**

A screen shot of a computer

Description automatically generated

**Results:**  
A screenshot of a computer screen

Description automatically generated



A screenshot of a computer

Description automatically generated

**Fifth anomaly detected (receipt ID:104312 and 118551):**

A screenshot of a computer

Description automatically generated

**Receipt ID: 104312 shares two different customer IDs. The customer names are Brandon Bennet and Justin Gray.**

A screenshot of a computer

Description automatically generated

**Receipt ID: 118551 shares two different customer IDs. The customer names are Christina Ross and Mia Clark.**

A screenshot of a computer

Description automatically generated

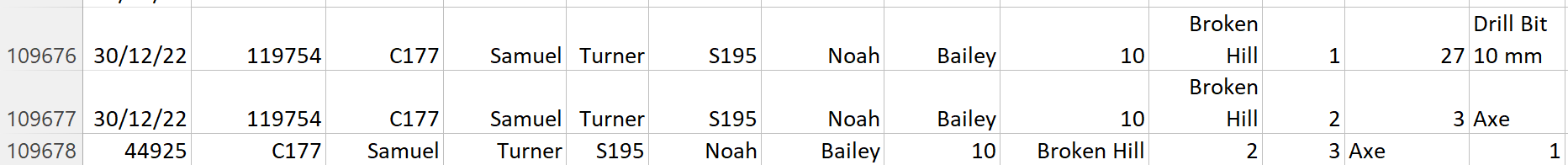
### 4.2.1 Solutions for cleaning the Data

Row: 109678 has multiple errors apparently because the values are shifted by 1 due to it missing a sales date, and receipt IDs: 104312 and 118551 share more than one customer ID.

**Solutions to fix the erroneous row 109678 anomalies:**

1. Omit the entire row.
2. Setting the row 109678 sales date and receipt ID that follows the receipt ID: 119754 and treating it as another transaction row in that receipt since the values of row 109678 have the same customer, staff, and office as receipt: 119754.
3. Placing the latest date in the row and shifting the values down.

We have decided to choose option two as we do not want to omit any row that might affect the analysis. The erroneous row might be insignificant, but we are aiming for precision. We are assuming that the erroneous row 109678 is associated with the previous row as the important values inside row 109677 are almost identical to row 109678. This way, we set row 109678 sales date and receipt ID to follow receipt ID 119754 to become the third transaction row for it, afterwards we will shift the values of each column by one as it appears that the values are shifted by one. This would fix the cause of all errors in row 109678.



**Solutions to fix the receipt ID belonging to two customers.**

1. Omit the row where the receipt ID appears in another customer ID.
2. Supplement an alphabet to the duplicated receipt ID.

We have decided to go with option 2 as previously mentioned we are aiming for precision. Supplementing an alphabet to the duplicated receipt ID would allow each customer to have a unique receipt ID while avoiding any loss of data that can significantly impact the analysis.

## 4.3 ETL Processes

This is the data flow of the ETL in SSIS. The files would be imported as a CSV file and will flow through the conditional split where it checks for any empty values and splits them between empty row total and no empty row total flow. The empty row total flow will join with no empty row total flow after it is cleaned to go through another script to fix the duplicated receipt IDS. Finally, the data would be passed to the SQL as “Clean Data.”

A screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generated

**Conditional split to check whether the row total is empty or not empty.**A screenshot of a computer

Description automatically generated

**Cleaning script for the erroneous row 109678 to change the structure of its values and to change its receipt ID to follow receipt ID 119754 and date.**

A screen shot of a computer program

Description automatically generated

**Remove duplication of receipt ID script to add a “D” to duplicated receipt IDs for rows: 104312 and 118551.**

A computer screen with text

Description automatically generated

A computer screen shot of a program code

Description automatically generated

## 4.4 Assumptions On staffID

In the provided Excel sheet, there were some staff sharing the same name and surname, We have decided to assume that each staff ID shares a single individual and it is very possible for an organisation to have two staff that share the same first name and surname.

**Queries and result that shows different staff id sharing the same name “Paul Guiterrez.”**

A screenshot of a computer

Description automatically generated



A table of names

Description automatically generated with medium confidence

Queries and result shows different staff id sharing the same name “Isabella Rogers.” For staff ID S24 the office location is located at “Wagga Wagga” and for staff ID S157 office location is located at “Dubbo”

A table with numbers and letters

Description automatically generated

A table with numbers and letters

Description automatically generated

## 4.5 Implementation of a Dashboard

In Power BI, a few pages were created in order to get the answers for Smotherington.

1. An overview of “Bits&Bobs”  
   A screenshot of a computer

   Description automatically generated
2. A dedicated staff KPI page with branch slicing  
   A screenshot of a graph

   Description automatically generated
3. A dedicated Item KPI page with branch slicing  
   A screenshot of a graph

   Description automatically generated
4. A visualisation of the difference in Incentive costs incurred by “Bits&Bobs”  
   A screenshot of a phone

   Description automatically generated
5. A branch sales forecast with branch slicing  
   A graph with lines and dots

   Description automatically generated with medium confidence

## 4.6 Base Analysis

**Using the dashboard and the Power BI functions we were able to address most of Smotherington’s concerns.**

### **4.6.1 Sales staff KPI**

**Top Salesperson(Revenue):**

1. **Overall: Savannah Jones**
2. **Broken Hill: Emma Gray**
3. **Dubbo: Zoe Barnes**
4. **Lismore: Samuel Anderson**
5. **Maitland: Kaitlyn Nguyen**
6. **New Castle: Stephanie Watson**
7. **Port Macquarie: Jenna Cox**
8. **Tammworth: Emma Gutierrez**
9. **Wagga Wagga: Mia Foster**
10. **West Wyalong: Austin Morris**
11. **Wollongong: Savannah Jones**

**Top Salesperson(Qty):**

1. **Overall: Samuel Anderson**
2. **Broken Hill: Carlos Sanders**
3. **Dubbo: Kaitlyn Scott**
4. **Lismore: Samuel Anderson**
5. **Maitland: Anna Hail**
6. **New Castle: Leah Harris**
7. **Port Macquarie: Jenna Cox**
8. **Tammworth: Emma Gutierrez**
9. **Wagga Wagga: Isaiah Powell**
10. **West Wyalong: Austin Morris**
11. **Wollongong: Savannah Jones**

**Top Salesperson(Transactions):**

1. **Overall: Samuel Anderson**
2. **Broken Hill: Carlos Sanders**
3. **Dubbo: Zoe Barnes**
4. **Lismore: Samuel Anderson**
5. **Maitland: Anna Hail**
6. **New Castle: Leah Harris**
7. **Port Macquarie: Jenna Cox**
8. **Tammworth: Emma Gutierrez**
9. **Wagga Wagga: Mia Foster**
10. **West Wyalong: Austin Morris**
11. **Wollongong: Savannah Jones**

### **4.6.2 Item KPIs**

**The items performing the best and the worst overall are:  
Best:**

1. **Cordless Drill kit**
2. **Socket Set**
3. **Grinder**

**Worst:**

1. **Drill bit 2mm**
2. **Drill bit 3mm**
3. **Drill bit 4mm**

### **4.6.3 Branch KPIs**

A screenshot of a graph

Description automatically generated

Overall, the Branch sales are reflective of the number of sales staffs assigned to those branches. An example of which is Wagga Wagga housing 32 of the 200 total sales staff. Wollongong, with the lowest sales staff also generated the least revenue. However, the overall best salesperson: Savannah Jones hails from the Wollongong Office. This could be attributed to lesser competition in the area but could also indicate that the branches overall performance is directly proportional to the number of sales staffs.

If the number of staff is indeed directly proportional to the revenue generated by the branches we can crudely infer which branches are making use of their resources the best.

1. **Wagga Wagga:**

Revenue: ~$1.83 million

Staff: 32

Revenue per Staff = $1.83 million / 32 = $57,187.50 per staff

1. **Maitland:**

Revenue: $1.25 million

Staff: 22

Revenue per Staff = $1.25 million / 22 = $56,818.18 per staff

1. **Broken Hill:**

Revenue: ~$1.22 million

Staff: 22

Revenue per Staff = $1.22 million / 22 = $55,454.55 per staff

1. **Lismore:**

Revenue: ~$1.13 million

Staff: 20

Revenue per Staff = $1.13 million / 20 = $56,500 per staff

1. **Newcastle:**

Revenue: ~$1.12 million

Staff: 20

Revenue per Staff = $1.12 million / 20 = $56,000 per staff

1. **Port Macquarie:**

Revenue: ~$1.11 million

Staff: 19

Revenue per Staff = $1.11 million / 19 = $58,421.05 per staff

1. **Tamworth:**

Revenue: ~$1.02 million

Staff: 17

Revenue per Staff = $1.02 million / 17 = $60,000 per staff

1. **West Wyalong:**

Revenue: ~$1.01 million

Staff: 16

Revenue per Staff = $1.01 million / 16 = $63,125 per staff

1. **Dubbo:**

Revenue: ~$944,000

Staff: 17

Revenue per Staff = $944,000 / 17 = $55,529.41 per staff

1. **Wollongong:**

Revenue: ~$902,000

Staff: 15

Revenue per Staff = $902,000 / 15 = $60,133.33 per staff

With this the least efficient branch would be **Broken Hill**. This could be that the staff are finding it more difficult to source demand for the products in Broken Hill. However this is not the only metric in which to measure regarding the censure of a branches operations.

**Forecasting**   
Forecasting could prove valuable to project the businesses health in those branches. The data “Bits&Bobs” provided us, give us very little historical data to extract a pattern, these projections ay be less accurate as such. However, we still believe that it is worth examining and have shared the charts below:

**Broken Hill  
A graph on a white background

Description automatically generated**

The trend Line indicates minor to no growth.

**Dubbo**A graph with a line going up

Description automatically generatedTrending upwards

**Lismore  
A graph with a line going up

Description automatically generated**Trendline indicates a downward trend but the projections seem to suggest high end year sales

**Maitland**  
A graph with a line going up

Description automatically generated  
Trendline indicates a downward trend

**New Castle  
A graph showing a line of a graph

Description automatically generated with medium confidence**The trend Line indicates minor to slight growth.

**Port Macquarie  
A graph with a blue line

Description automatically generated**Trend line indicates larger growth, projections seem marginal

**Tammworth**  
**A graph with a line going up

Description automatically generated**Trendline indicates downward trajectory for the branch

**Wagga Wagga  
A graph showing a line and a line

Description automatically generated with medium confidence**Trendline indicates downward trend, but wagga wagga is the most profitable branch and the projections indicate stability

**West WyalongA graph showing a line and a graph showing a line

Description automatically generated**West Wyalong was also indicated as the most efficient branch in terms of profit per staff member. Trendline indicates steady growth. Projections dictate otherwise.  
  
**Wollongong** **A graph with a line and a black box

Description automatically generated with medium confidence**Wollongong trends downwards.

### 4.6.4 Incentive Measures

As requested by Smotherington, based on the previous years data, we have calculated the cost incurred by “Bits&Bob” in exercising sales initiatives for customers. The comparison between the recently implemented 5% discount and the proposed customer point system.

A screenshot of a phone

Description automatically generated

At 250 points to a dollar, the organization saves an estimated $421,000 per year on the incentive change alone.

# 