

Yeanie Lamin

{Data Analytics Portfolio}

Table of Contents

03

Meet Yeanie Lamin

06

**Does *Alpha-ketoglutarate*
Affect *Nessonia vitripennis*?**

04

Skills-based *Curriculum Vitae*

13

**The Customer Satisfaction
Project**

05

**The value of data for
organisations**

17

**The Brewery Operations
Project**



Hi, I'm Yeanie.

Welcome to my data analytics portfolio.

I am a passionate perspective data analyst with a strong foundation in data wrangling, data visualization, and statistical analysis. Over the last several years I have developed the ability to transform raw data into insights that have enhanced my understanding primarily of science.

I am a recent graduate and hold a Bachelor of Science with Honours in Medical Physiology from the University of Leicester. My degree has enhanced my understanding of biology & medicine, scientific research methodologies, and statistical analysis. Through my degree, I was able to begin my programming journey with R and GraphPad Prism, two highly powerful tools for academic research, and performing statistical tests.

In the last couple of months, I have embarked on upskilling my data knowledge with Cambridge Sparks, where I received a Level 3 Data Citizen Skills Bootcamp. This program has enhanced my skills in Microsoft Excel, Power Query (also known as Get & Transform), Microsoft Power BI, and a brief introduction to Tableau. The bootcamp also introduced me to SQL and Python, providing me with the foundation to build on these skills further.

To date, I have worked on a diverse array of projects that have exercised my ability to clean and wrangle data, visualise data, build data models, perform regression curves, and perform statistical tests. I am eager to expand my toolkit and knowledge of data analytics, and understanding of business practices and principles.

This portfolio showcases a few of the projects I have worked on, highlighting my ability to solve complex problems, communicate my findings effectively, and deliver results that make a tangible impact.

I invite you to explore my work and learn about my approach to analysing data.



www.linkedin.com/in/yeanie-lamin-404206162



yeaniel30@gmail.com

CV

Personal Details



London, UK



yeaniel30@gmail.com



[linkedin.com/in/yeanie-lamin-404206162](https://www.linkedin.com/in/yeanie-lamin-404206162)



<https://github.com/yeaniel/yeaniel.git>

Technical Skills

Microsoft Excel

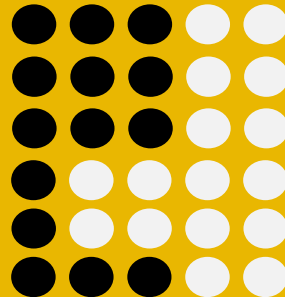
Microsoft Power BI

R

SQL

Python

GraphPad Prism



Education & Certificates

Cambridge Sparks

Level 3 Data Skills Bootcamp

February 2024 – April 2024

Remote, UK

University of Leicester

Bachelor of Science (Honours) in Medical Physiology, 2:1

September 2019 – June 2023

Leicester, UK

R

- I used R to analyse data for my research project where I investigated the effects of a compound Alpha-ketoglutarate on the movement of parasitic wasps.
- I used the tidyverse and lubridate packages to wrangle and clean my data. I used the ggplot2 package to create boxplots with my data. I used the base package to perform, a Wilcoxon test and create a Poisson regression model.

Microsoft Excel

- I used Microsoft Excel as an administrator to input property data into an integrated database.
- I use functions such as Vlookup, Xlookup, and index matching to cross reference data from different workbooks.
- I've used functions such as sumifs, and averageifs to make calculations based on select criteria.
- I've used pivot tables to efficiently make calculations and pivot charts with slicers.

Microsoft Power BI

- I used Microsoft Power BI to create interactive dashboards to easily and efficiently communicate data.
- I used get & transform (also known as power query) to efficiently clean data and create calculated columns.

The value of data for organisations

Data can be defined as information that is stored or used by a computer. Data can be numerical, text-based, date and time-based, or Boolean.

Data follows 7 principles, known as the “7 Vs”. Today, there is a large *volume* of data that can be gathered quicker than ever at high *velocity*. Data can be present in a *variety* of formats, from a wide range of sources including websites, financial reports, social media, government polls, healthcare, and bank statements. Data is also *variable* and can have different distribution patterns or frequencies. Data has flexible *veracity*, which means that it can differ in accuracy, precision, and trustworthiness. Data can also be *visualised* to communicate information about what data means. Finally, data has *value*, as different types of data are important to different types of organisations.

Organisations gather data to better understand trends. In doing so, organisations can market products or services more effectively and increase their revenue or impact in an industry. A pharmaceutical company like GSK would find scientific research highly valuable to create products to treat specific disorders. Whereas a social media company such as Meta would find the type of content people post and the way people interact with content valuable to increase the number of users of its platforms. A financial services provider such as JP Morgan would find financial transactions valuable to determine the best stocks and bonds to buy and sell.

The value of data has increased with the wide availability of data, allowing organisations to respond to trends more accurately. Therefore, organisations can increase their revenue, and significance in a specific industry.



A faint, light green background image of a Nasonia vitripennis fly, showing its head, thorax, and legs in profile, facing left.

Can α -ketoglutarate improve the activity levels of *Nasonia vitripennis*?

Background

This project is based on my university research experiment, which investigated the effects of alpha-ketoglutarate (AKG) on the movement of *Nasonia vitripennis*. The aim of the study was to determine if AKG could increase the healthspan of *N. vitripennis*. AKG is a naturally occurring compound in our bodies, that forms part of the Krebs cycle. Various studies have shown that dietary AKG can be associated with an increase in healthspan.

My study became of interest due to the human ageing process, where ageing can be defined as a gradual decline in physiological function, eventually leading to death. In our modern world, the life expectancy of individuals in many countries is increasing although many of individuals are not leading healthier lives. This is associated with a variety of factors such as being female, being from a lower socioeconomic background or obesity.

The Background

The Effect of *Alpha-ketoglutarate* on *Nasonia vitripennis*

Problems with an ageing population

Many countries around the world have an ageing population, due to an increase in life expectancy and low birth rates. Although in many industrialised nations, people are not living healthier lives despite living longer. Studies have shown that approximately 16 – 20% of later life is spent in ill health. Additionally, these studies have found that those who are female, from a lower socioeconomic background or obese are at risk of ill health in the latter stages of life.

Nasonia vitripennis

Nasonia vitripennis is a parasitic wasp that uses the larvae of Blowflies and similar species as a host. *N. vitripennis* was used as an animal model due to the species experiencing DNA methylation at specific motifs in their cytosine residues. DNA methylation is the transfer of a methyl group from S-adenyl methionine to a DNA sequence. This process typically occurs at specific points in DNA including cytosine and adenosine residues. DNA methylation is thought to be an epigenetic cause of ageing in humans and some other multicellular organisms. Unlike *N. vitripennis*, other distantly related species such as bees and ants do not experience this type of DNA methylation.

Sample

My study had two groups of male and female *N. vitripennis*. The experimental group were fed AKG and sucrose solution. Whilst the control group was just fed sucrose solution.

Experimental Design

My study measured five aspects of activity in the wasps. This included the percentage of time spent moving and the number of:

- Jumps
- Hops – defined by a distance of less than 2cm.
- Short flights – defined by a distance of 2 – 4cm.
- Long flights – defined by a distance of more than 4cm.

Analysing My Data

The Effect of *Alpha-ketoglutarate* on *Nasonia vitripennis*

Tools



R Studio (R code)

Sourcing my data

```
library(readxl)
jumps <- read_excel("~/Data Analyst Portfolio/AKG on Activity
Levels/jumps.xlsx")
## New names:
## • `` -> `...2`
```

Wrangling Data

Rename columns in the table

```
names(jumps)[2] <- "sex"
names(jumps)[3] <- "treatment"
names(jumps)[4] <- "jumps.no"
names(jumps)[5] <- "short.flight"
names(jumps)[6] <- "long.flight"
```

Changing the format of the vector - POSIXct to character in minutes and seconds

```
jumps$`Moving Time` <- as.character(jumps$`Moving Time`, format="%M:%S")
jumps$`Stopped Time` <- as.character(jumps$`Stopped Time`,
format="%M:%S")
```

Creating a new column in the table - time in seconds

```
library(lubridate)
##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union
jumps$Moving.Time <- as.numeric(ms(jumps$`Moving Time`))
jumps$Stopped.Time <- as.numeric(ms(jumps$`Stopped Time`))
```

Creating a new column in the table - percentage of time spent moving

```
jumps$percent <- (jumps$Moving.Time /
(jumps$Stopped.Time+jumps$Moving.Time))*100
```

Visualising my data

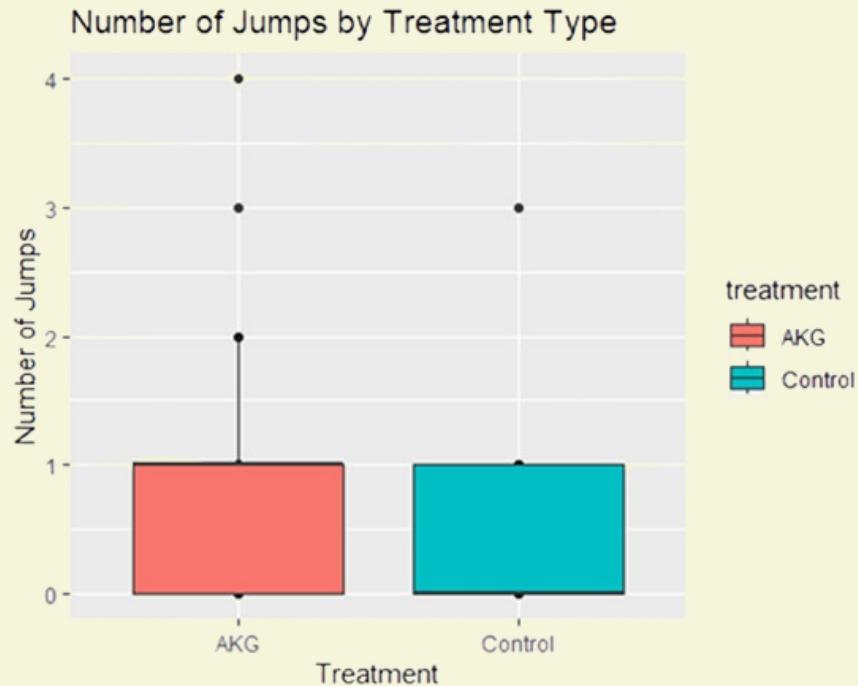
```
library(ggplot2)
```


The Results

The Effect of *Alpha-ketoglutarate* on *Nasonia vitripennis*

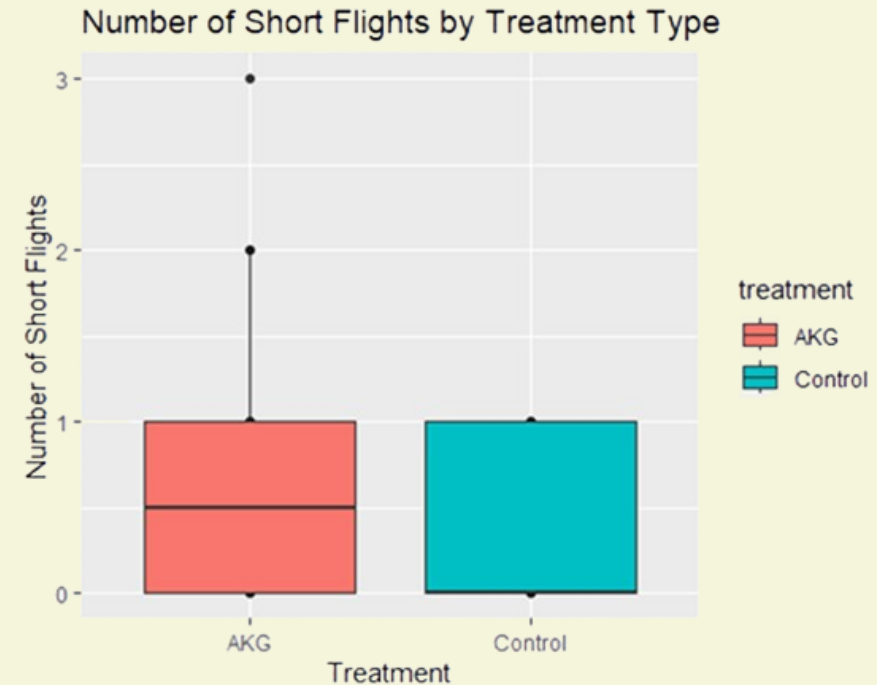
Boxplot for the number of jumps

```
ggplot(jumps, aes(x=treatment, y=jumps.no, fill=treatment)) +  
  geom_point() + geom_boxplot() +  
  xlab("Treatment") +  
  ylab("Number of Jumps") +  
  ggtitle("Number of Jumps by Treatment Type")
```



Boxplot for the number of short flights

```
ggplot(jumps, aes(x=treatment, y=short.flight, fill=treatment)) +  
  geom_point() + geom_boxplot() +  
  xlab("Treatment") +  
  ylab("Number of Short Flights") +  
  ggtitle("Number of Short Flights by Treatment Type")
```

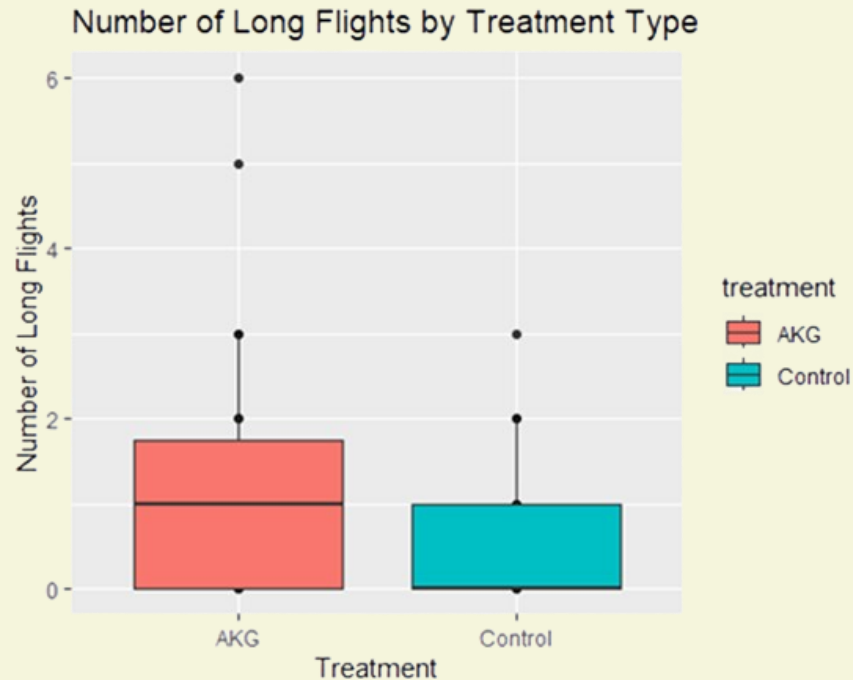


The Results

The Effect of *Alpha-ketoglutarate* on *Nasonia vitripennis*

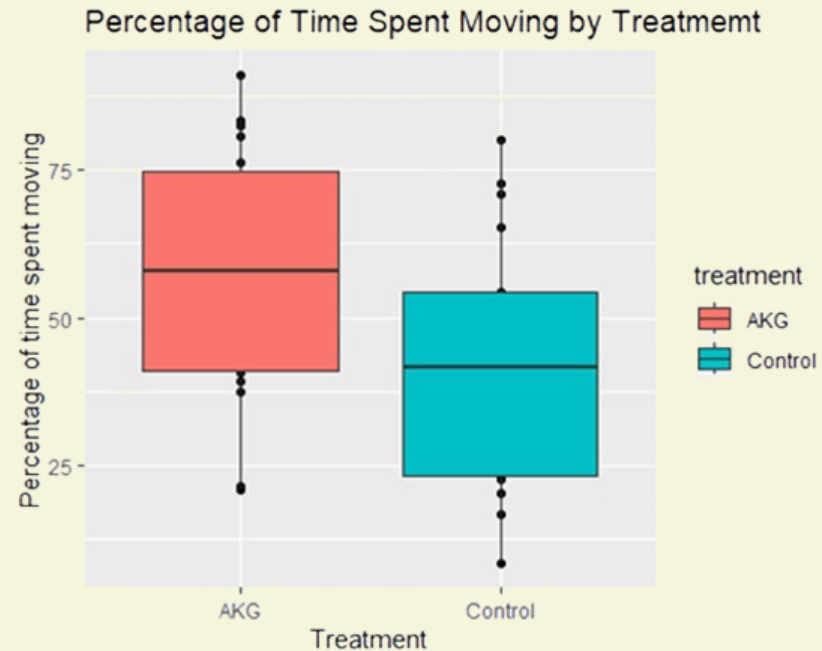
Boxplot for the number of long flights

```
ggplot(jumps, aes(x=treatment, y=long.flight, fill=treatment)) +  
  geom_point() + geom_boxplot() +  
  xlab("Treatment") +  
  ylab("Number of Long Flights") +  
  ggtitle("Number of Long Flights by Treatment Type")
```



Boxplot for the percentage of time spent moving

```
ggplot(jumps, aes(x=treatment, y=percent, fill=treatment)) +  
  geom_point() + geom_boxplot() +  
  xlab("Treatment") +  
  ylab("Percentage of time spent moving") +  
  ggtitle("Percentage of Time Spent Moving by Treatment")
```



Statistics

The Effect of *Alpha-ketoglutarate* on *Nasonia vitripennis*

Statistical tests

Wilcoxon tests for the number of jumps, short flights, and long flights

```
wilcox.test(formula = jumps.no ~ treatment, data = jumps, exact=FALSE)
##
## Wilcoxon rank sum test with continuity correction
##
## data: jumps.no by treatment
## W = 239.5, p-value = 0.09922
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(formula = long.flight ~ treatment, data = jumps,
exact=FALSE)
##
## Wilcoxon rank sum test with continuity correction
##
## data: long.flight by treatment
## W = 246.5, p-value = 0.07075
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(formula = short.flight ~ treatment, data = jumps,
exact=FALSE)
##
## Wilcoxon rank sum test with continuity correction
##
## data: short.flight by treatment
## W = 223.5, p-value = 0.2467
## alternative hypothesis: true location shift is not equal to 0
```

Statistics

The Effect of *Alpha-ketoglutarate* on *Nasonia vitripennis*

#Poisson regression

```
Poisson_model <- glm(jumps.no + short.flight + long.flight +
  percent ~ treatment,
  data = jumps, family = poisson(link= "log"))

## Warning in dpois(y, mu, log = TRUE): non-integer x = 37.500000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 74.333333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 72.166667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 44.762063
## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.833333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 45.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 72.166667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 23.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 48.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 41.166667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 67.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 21.333333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 44.500000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 8.500000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 24.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 58.333333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 31.833333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 66.166667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 23.333333
```

```
## Warning in dpois(y, mu, log = TRUE): non-integer x = 41.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 41.500000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 56.217391
## Warning in dpois(y, mu, log = TRUE): non-integer x = 88.794702
## Warning in dpois(y, mu, log = TRUE): non-integer x = 51.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 44.529750
## Warning in dpois(y, mu, log = TRUE): non-integer x = 37.628866
## Warning in dpois(y, mu, log = TRUE): non-integer x = 79.333333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 84.392027
## Warning in dpois(y, mu, log = TRUE): non-integer x = 87.695507
## Warning in dpois(y, mu, log = TRUE): non-integer x = 83.333333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 99.014975
## Warning in dpois(y, mu, log = TRUE): non-integer x = 74.787980
## Warning in dpois(y, mu, log = TRUE): non-integer x = 80.033278
## Warning in dpois(y, mu, log = TRUE): non-integer x = 16.666667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 52.059308
```

Printing my Poisson regression model

```
##
## Call:  glm(formula = jumps.no + short.flight + long.flight +
  percent ~
  treatment, family = poisson(link = "log"), data = jumps)
##
## Coefficients:
##      (Intercept)  treatmentControl
##           4.1065           -0.3407
##
## Degrees of Freedom: 38 Total (i.e. Null);  37 Residual
## Null Deviance:      419.8
## Residual Deviance: 363.2    AIC: Inf
```


The Customer Satisfaction Project



Tools



Microsoft
Power BI

About The Data

A dataset based on customer reviews on Google Maps of a group of restaurants in Turkey.

The dataset contains:

- Restaurant names
- Author names
 - Text
 - Photos
 - Rating
 - Rating category - considers *taste, menu, and indoor and outdoor atmosphere*.

Source: *Deniz Bilgin*. [Google Maps Restaurant Reviews \(kaggle.com\)](https://www.kaggle.com/denizbilgin/google-maps-restaurant-reviews)

Cleaning Process

The Customer Satisfaction Project

- Imported data into Power Query
- Renamed columns
- Replaced values

e.g.

outdoor_atmosphere →
outdoor atmosphere

Renaming columns

Replacing values

Table Data (from top screenshot):

	Business name	Author Name	Comments	Photo Path
1	Hacı'nın Yeri - Yigit Lokantasi	Gulsum Akar	We went to Marmaris with my wife for a holiday. We chose this restau...	dataset/taste...
2	Hacı'nın Yeri - Yigit Lokantasi	Oguzhan Cetin	During my holiday in Marmaris we ate here to fit the food. It's really g...	dataset/men...
3	Hacı'nın Yeri - Yigit Lokantasi	Yasin Kuyru	Prices are very affordable. The menu in the photo cost 108 liras. You h...	dataset/outdo...
4	Hacı'nın Yeri - Yigit Lokantasi	Orhan Kapu	Turkey's cheapest artisan restaurant and its food is delicious!	dataset/indoc...
5	Hacı'nın Yeri - Yigit Lokantasi	Ozgur Sati	I don't know what you will look for in terms of price performance poin...	dataset/men...
6	Hacı'nın Yeri - Yigit Lokantasi	Arda Karaca	Generally good.	dataset/indoc...
7	Hacı'nın Yeri - Yigit Lokantasi	A'rem Eren	What you see is 125 TL in total. It's a pretty convenient place. W...	dataset/indoc...
8	Hacı'nın Yeri - Yigit Lokantasi	Nadia Salim	Delicious food at rock bottom prices. Friendly staff; Highly recom...	dataset/indoc...
9	Hacı'nın Yeri - Yigit Lokantasi	Mehmet Eser	Every time I go, I still experience the amazement I experienced y...	dataset/indoc...
10	Hacı'nın Yeri - Yigit Lokantasi	Celal Ozer	The most 1/5 of all businesses I've seen.	dataset/indoc...
11	Hacı'nın Yeri - Yigit Lokantasi	Onur Celebi	The food was just like the ones my mother made. Prices were re...	dataset/indoc...
12	Pizza Fellas	Kadir Tasci	The ambiance of the place is nice; when we went; there was a lo...	dataset/indoor...
13	Pizza Fellas	Serkan Colak	Flavor : It has nothing but dough taste. Labor : Not befitting a to...	dataset/taste/p...
14	Pizza Fellas	Selim Cun	It is in the center; easy to reach; close to the parking lot; the pric...	dataset/indoor...
15	Pizza Fellas	Zeki Celik	They make delicious pizza with nice ingredients on thin dough. Si...	dataset/indoor...
16	Pizza Fellas	Mustafa Caliskan	It's in a nice place but it's very small and their pizza is not that go...	dataset/outdoor...
17	Pizza Fellas	Deniz Bilgin	I strongly recommend. If I come to the taste part; it was very enj...	dataset/taste/p...
18	Pizza Fellas	Cihan Tas	It is very tasty; similar to Italian pizzas. Roast beef and mohair 18...	dataset/outdoor...

Cleaning Process

The Customer Satisfaction Project

Adding conditional a column

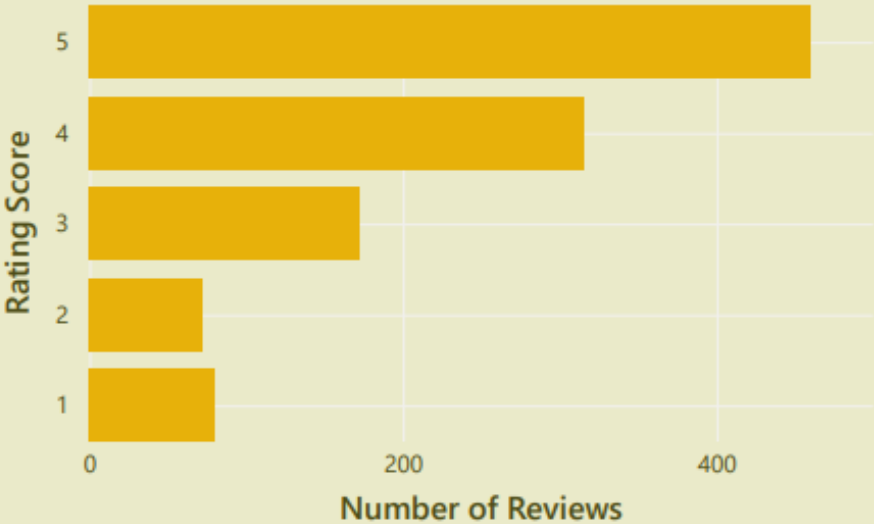
Removing a column

- Added a conditional column
- Review outcomes to suggest if comments were overall:

- Positive
- Negative
- Okay

- Removed columns
- Removed a photo path column.

Rating Score by Number of Reviews

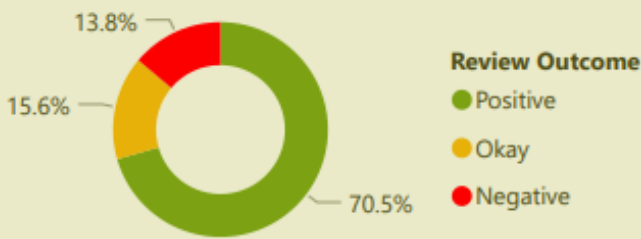


Restaurants

Abidin Tantuni

Ahsap Iskender

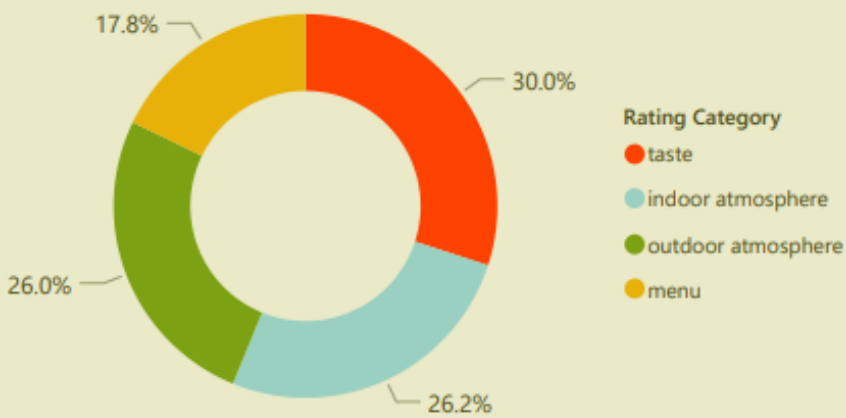
Customer Satisfaction



Average Score



Percentage of Reviews by Rating Category



Business name	Comments
Zula	Burgers are great hot dogs do. It is very interesting that french fries are 114 TL.
Zula	Burgers are super. I've eaten hamburgers in many places; but I've never come across anything like t meal with friendly service in a small but cozy place. I strongly recommend.
Zula	Hamburger is really delicious. The fries are normal.
Zula	In general; I was satisfied; since the hamburger is a little small; it will be more satisfying if the one w grammage is chosen.
Zula	It is delicious in everything with its sauce and bread.
Zula	It was one of the best burgers I've ever eaten. The only downside is that the portion is a little small. it is definitely a place to experience.
Zula	Surprised location; but nice clean venue. The employees are friendly; I don't want to comment on tl were too high..
Zula	The best hamburger restaurant in Istanbul. Cooking and sauce is very good.

The Brewery Operations Project

Tools



Microsoft
Excel

About the data

A dataset on craft beer production of a company in Bangalore, India which consists of sales, beer quality, and brewing operations parameters.

Source

Ankur Napa. [Brewery Operations and Market Analysis Dataset \(kaggle.com\)](https://www.kaggle.com/ankurnapa/brewery-operations-and-market-analysis-dataset)



The Sales Problem

The Brewery Operations Project



A craft brewery company in India saw sales decrease in January by **\$30 million** a day.



The craft brewery company wanted to increase its sales in February back up to \$70 million a day minimum.



The craft brewery company wanted to increase sales by improving the efficiency of its brewery operations.



Sales Breakdown

The Brewery Operations Project



Highest sales

Jayanagar, Bangalore
\$107 million



Lowest sales

Electronic City, Bangalore
\$102 million



Most sold beverage

Stout
\$132.2 million



Least sold beverage

Wheat Beer
\$129.8 million



Most sold stock keeping unit

Cans
\$264.0 million



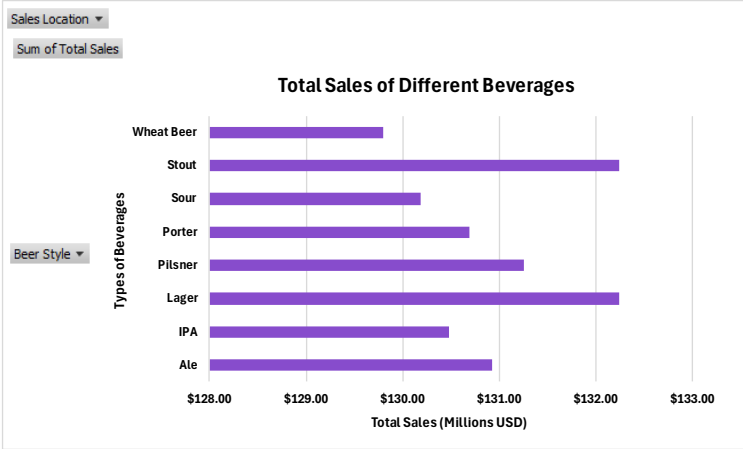
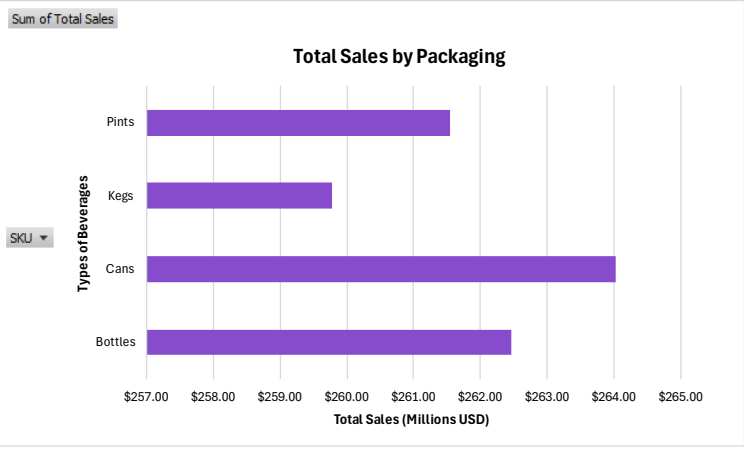
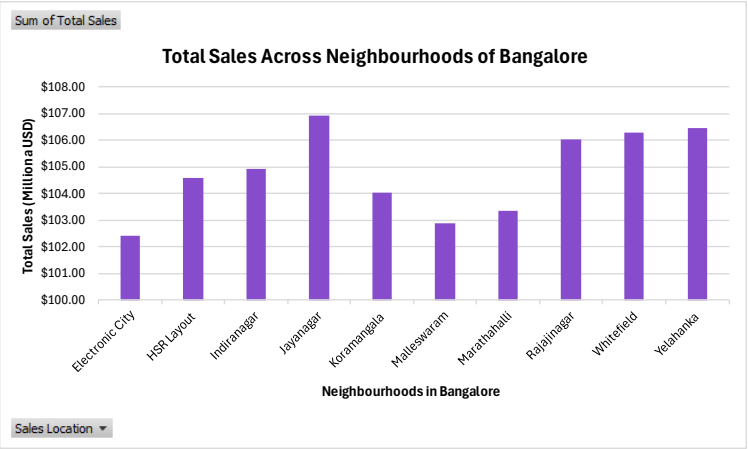
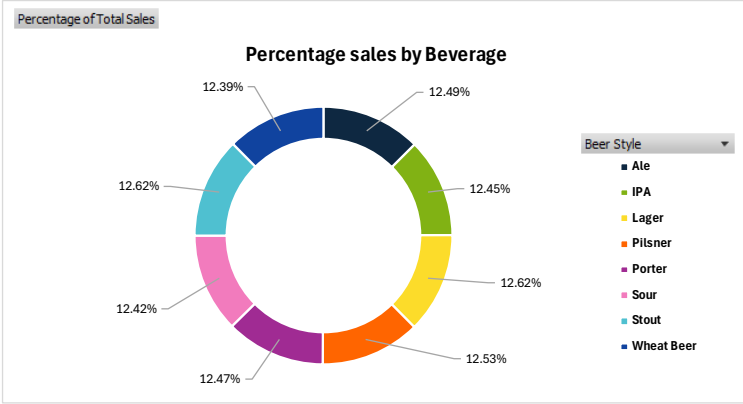
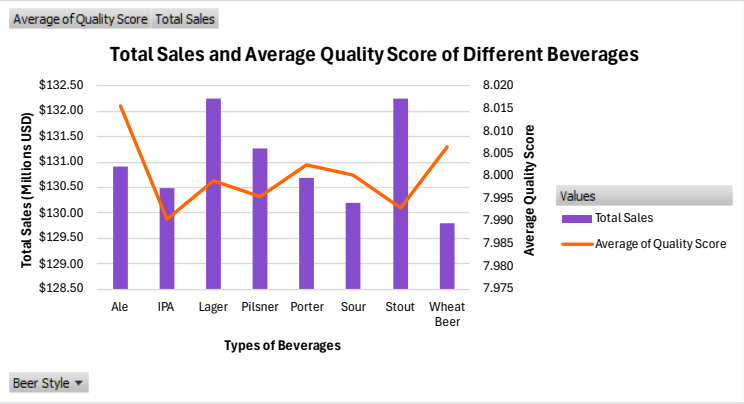
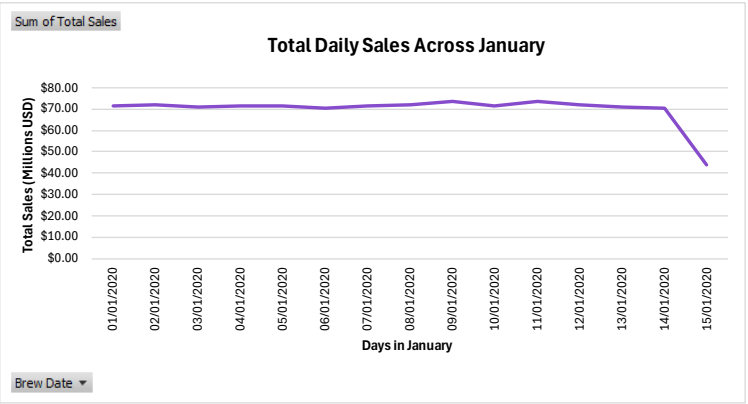
Least sold stock keeping unit

Kegs
\$259.8 million



Sales Breakdown

The Brewery Operations Project




- Sales Location
- Electronic City
 - HSR Layout
 - Indiranagar
 - Jayanagar
 - Koramangala
 - Malleswaram
 - Marathahalli
 - Rajajinagar
 - Whitefield
 - Yelahanka


Operations Breakdown

The Brewery Operations Project







SKU With The Highest Quality Score
Bottles
8.0099/10.0000



Beverage With The Lowest Quality Score
IPA
7.9905/10.0000
*Stout – close second




Ingredients ratio
All beverages
Similar ratio of water: grain: hops




SKU With The Lowest Quality Score
Pints
7.9910/10.0000
*Cans – close second



Most produced SKU
Cans
69.4 million litres
*Kegs – close second




Beverage With The Most Wastage
Stout & Larger
Highest volume loss during fermentation, brewing and bottling & kegging



Beverage With The Highest Quality Score
Ale
8.0150/10.0000



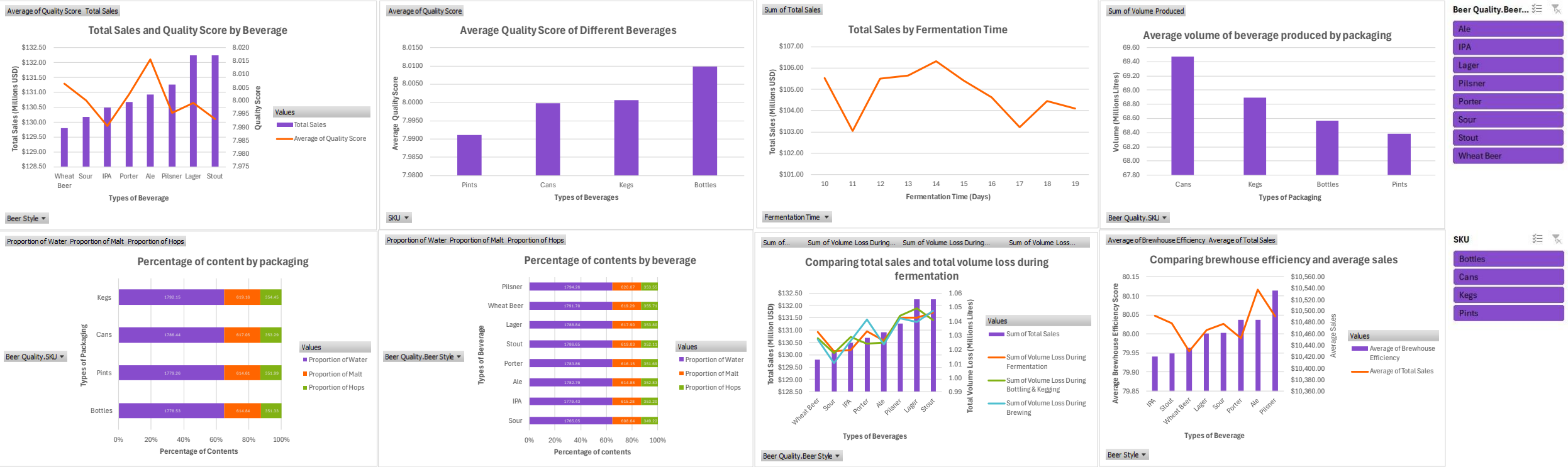
Least produced SKU
Pints
68.3 million litres



Beverage With The Least Wastage
Sour
Lowest volume loss during fermentation, brewing and bottling & kegging

Operations Breakdown

The Brewery Operations Project



Sum of... | Sum of Volume Loss During... | Sum of Volume Loss During... | Sum of Volume Loss During...

Comparing total sales and total volume loss during fermentation

Types of Beverages	Total Sales (Million USD)	Sum of Volume Loss During Fermentation	Sum of Volume Loss During Bottling & Kegging	Sum of Volume Loss During Brewing
Wheat Beer	129.50	0.01	0.01	0.01
Sour	130.00	0.01	0.01	0.01
IPA	130.50	0.01	0.01	0.01
Porter	131.00	0.01	0.01	0.01
Ale	131.50	0.01	0.01	0.01
Pilsner	132.00	0.01	0.01	0.01
Lager	132.50	0.01	0.01	0.01
Stout	133.00	0.01	0.01	0.01

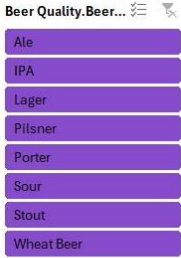
Beer Quality.Beer Style ▾

Average of Brewhouse Efficiency | Average of Total Sales

Comparing brewhouse efficiency and average sales

Types of Beverage	Average Brewhouse Efficiency Score	Average of Total Sales
IPA	79.95	10,500.00
Stout	79.95	10,480.00
Wheat Beer	79.95	10,460.00
Lager	79.95	10,440.00
Sour	79.95	10,420.00
Porter	79.95	10,400.00
Ale	79.95	10,380.00
Pilsner	79.95	10,360.00

Beer Style ▾



Conclusion
Beer quality is not the main factor that drives sales.
Produce less kegs.
Produce less wheat beer, to save costs.

The Cleaning Process

The Brewery Operations Project

File Home Insert Page Layout Formulas Data Review View Automate Help Power Query

Get Data Refresh Properties

From File From Excel Workbook From Text/CSV From XML From JSON From PDF From Folder From SharePoint Folder

From Text/CSV Import from file, From Web, From Database, From Azure, From Power Platform, From Online Services, From Other Sources, Combine Queries

Launch Power Query Editor... Data Source Settings... Query Options

Table.RenameColumns(#"Changed Type",{"Batch ID", "Batch ID"}, {"Brew Date", "Brew Date"}, {"Beer Style", "Beer Style"}, {"SKU", "SKU"}, {"Location", "Location"}),

	Batch ID	Brew Date	Beer Style	SKU	Location
1	7870796	01/01/2020 00:00:19	Wheat Beer	Keys	Whitefield
2	9810411	01/01/2020 00:00:31	Sour	Keys	Whitefield
3	2623342	01/01/2020 00:00:40	Wheat Beer	Keys	Malleswaram
4	8114651	01/01/2020 00:01:37	Ale	Keys	Rajajinagar
5	4579587	01/01/2020 00:01:43	Stout	Cans	Marathahalli
6	8715759	01/01/2020 00:01:48	Ale	Keys	Whitefield
7	6441292	01/01/2020 00:01:49	Lager	Pints	Electronic City
8	8843420	01/01/2020 00:01:51	Wheat Beer		
9	5713096	01/01/2020 00:02:16	Stout		
10	8178852	01/01/2020 00:02:32	Stout		
11	5607233	01/01/2020 00:02:36	Pilsner		
12	3699698	01/01/2020 00:02:55	IPA		
13	4437940	01/01/2020 00:03:04	Sour		
14	6167789	01/01/2020 00:03:09	Ale		
15	9487317	01/01/2020 00:03:12	Porter		
16	6902208	01/01/2020 00:03:17	Pilsner		
17	1816588	01/01/2020 00:03:22	Sour		
18	4910947	01/01/2020 00:03:26	Porter		

20 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

- Imported the data into Power Query
- Renamed columns
- Split 'Brew Date' column by a space delimiter into two columns – 'Brew date' and 'Time of brew'

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Space

Split at

☐ Left-most delimiter

☐ Right-most delimiter

☒ Each occurrence of the delimiter

Advanced options

Quote Character

"

☐ Split using special characters

Insert special character

Brew Date	Time of Brew
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
01/01/2020	00:00:19
01/01/2020	00:00:31
01/01/2020	00:00:40
01/01/2020	00:01:37
01/01/2020	00:01:43
01/01/2020	00:01:48
01/01/2020	00:01:49
01/01/2020	00:01:51
01/01/2020	00:02:16
01/01/2020	00:02:32
01/01/2020	00:02:36
01/01/2020	00:02:55
01/01/2020	00:03:04
01/01/2020	00:03:09
01/01/2020	00:03:12
01/01/2020	00:03:17
01/01/2020	00:03:22
01/01/2020	00:03:26

The Cleaning Process

The Brewery Operations Project

1

2

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Colon

Split at

☐ Left-most delimiter

☐ Right-most delimiter

☒ Each occurrence of the delimiter

Advanced options

Quote Character

"

☐ Split using special characters

Insert special character

3

Table.RenameColumns(#"Changed Type3",{"Ingredient Ratio.1","Proportion of Water", "Ingredient Ratio.2", "Proportion of Malt", "Ingredient Ratio.3", "Proportion of Hops" })

7

Keep Top Rows

Specify how many rows to keep.

Number of rows

100000

OK

Cancel

Separated the 'ingredients ratio' by a colon delimiter into three columns.

Renamed the tree columns and converted them into a percentage.

Once I finished cleaning the data, I limited the number of rows to 100,000 due to the dataset being voluminous.

Custom Column

Add a column that is computed from the other columns.

New column name

Percentage of Water

Custom column formula

$$= \left(\frac{([Volume\ Produced] * [Proportion\ of\ Water])}{([Proportion\ of\ Water] + [Proportion\ of\ Malt] + [Proportion\ of\ Hops])} \right) * 100$$

Available columns

Alcohol Content
Bitterness
Color
Proportion of Water
Proportion of Malt
Proportion of Hops
Volume Produced
Total Sales

<< Insert

Learn about Power Query formulas

Custom Column

Add a column that is computed from the other columns.

New column name

Percentage of Malt

Custom column formula

$$= \left(\frac{([Volume\ Produced] * [Proportion\ of\ Malt])}{([Proportion\ of\ Water] + [Proportion\ of\ Malt] + [Proportion\ of\ Hops])} \right) * 100$$

Available columns

Alcohol Content
Bitterness
Color
Proportion of Water
Proportion of Malt
Proportion of Hops
Volume Produced
Total Sales

6

Custom Column

Add a column that is computed from the other columns.

New column name

Percentage of Hops

Custom column formula

$$= \left(\frac{([Volume\ Produced] * [Proportion\ of\ Hops])}{([Proportion\ of\ Water] + [Proportion\ of\ Malt] + [Proportion\ of\ Hops])} \right) * 100$$

Available columns

Alcohol Content
Bitterness
Color
Proportion of Water
Proportion of Malt
Proportion of Hops
Volume Produced
Total Sales

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK

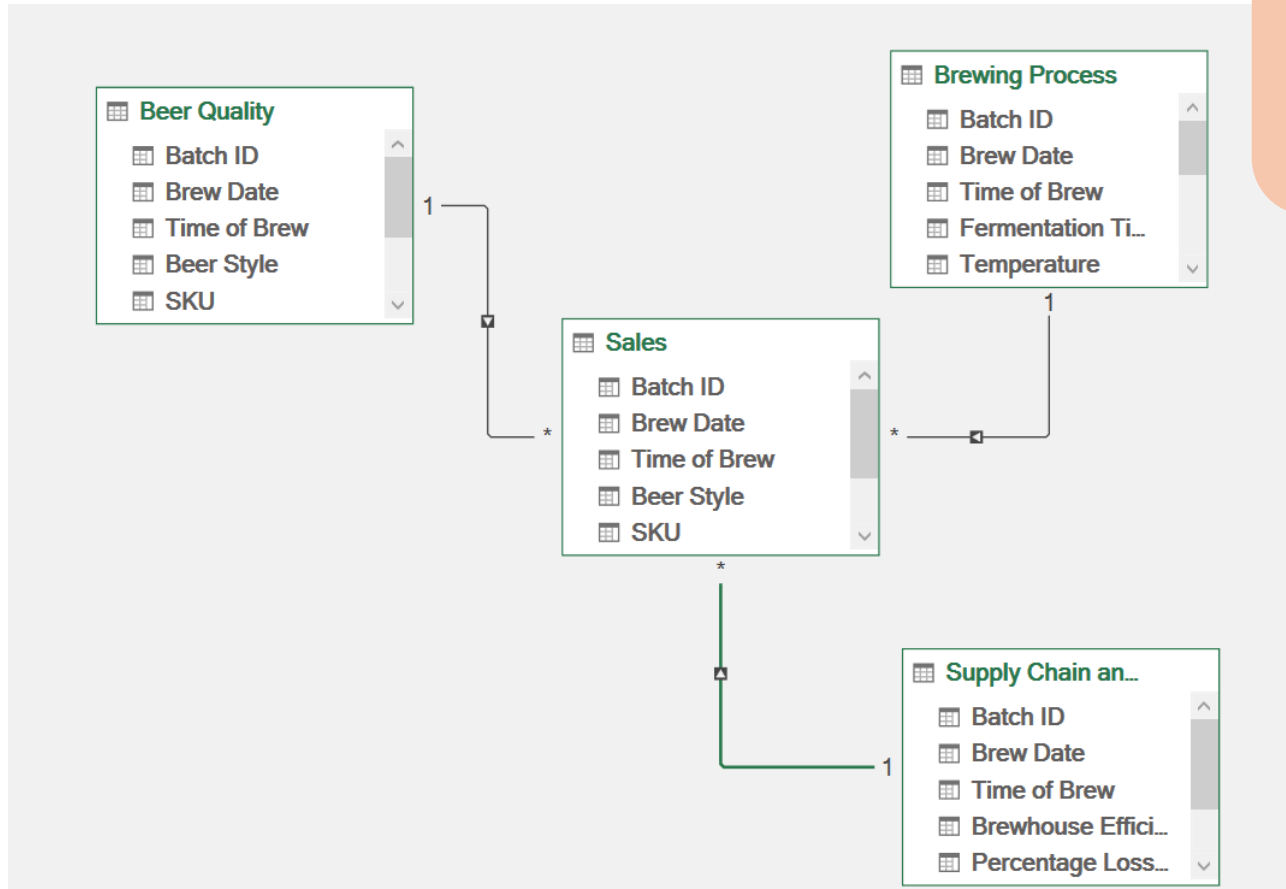
Cancel

024

The Data Model

The Brewery Operations Project

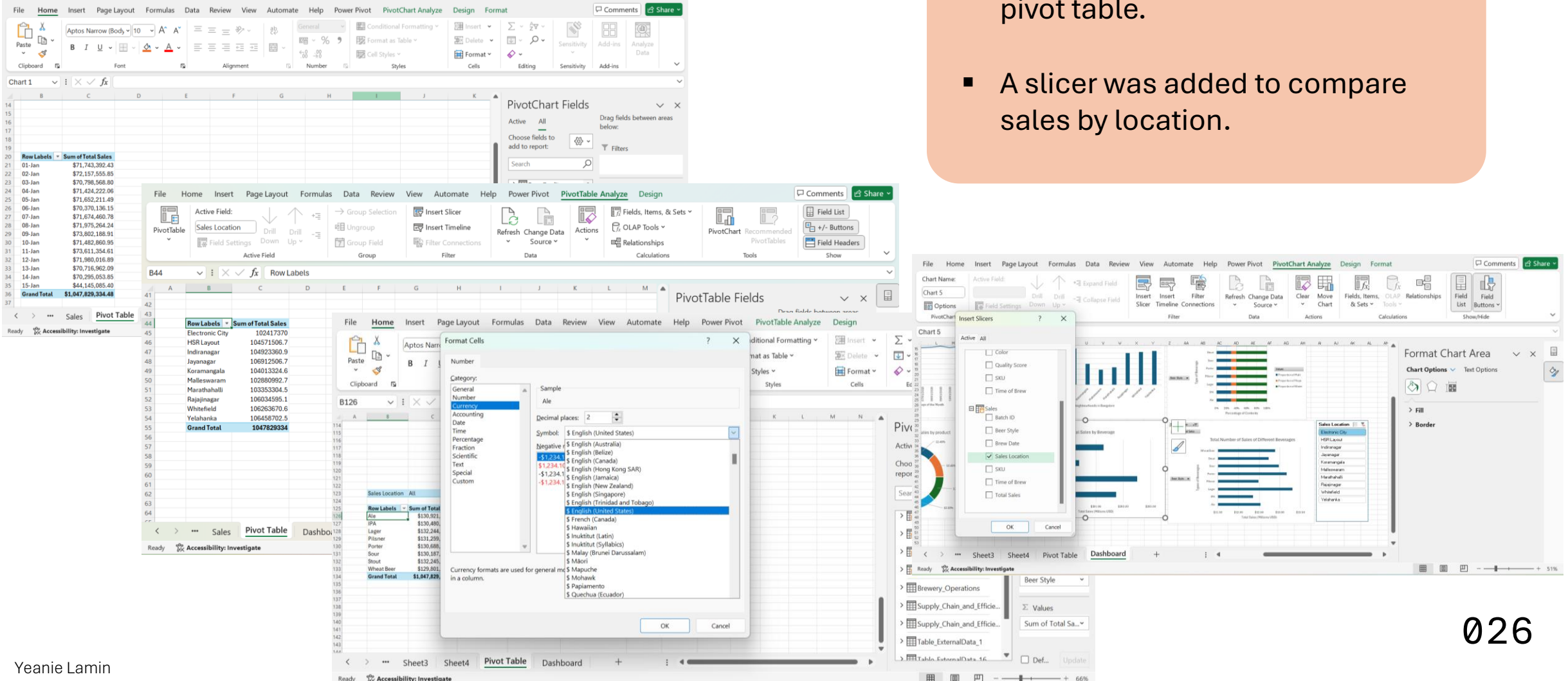
- Due to the large number of columns in the dataset, I broke the dataset down into 4 smaller datasets.
- I then created a data model.



Analysing the Data

The Brewery Operations Project

- Each graph was created using a pivot table.
- A slicer was added to compare sales by location.



Thank You!

