



# Yeanie Lamin

{Data Analytics Portfolio}

# Table of Contents

03

**Meet Yeanie Lamin**

05

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

04

**The value of data for  
organisations**

12

**The Customer Satisfaction  
Project**

16

**The Brewery Operations  
Project**



# Hi, I'm Yeanie.

I am a recent Medical Physiology graduate, with experience in R code and GraphPad Prism. I find that R is commonly used in scientific research, but less so in other industries.

I joined the Cambridge Sparks Skills Bootcamp to improve my use of Microsoft Excel and widen my knowledge of commonly used programs such as Python and SQL.



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



[yeaniel30@gmail.com](mailto:yeaniel30@gmail.com)



# 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 large, faint, light green image of a Nasonia vitripennis fly serves as the background for the slide. The fly is shown from a side profile, facing left, with its wings and legs visible.

# Can $\alpha$ -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

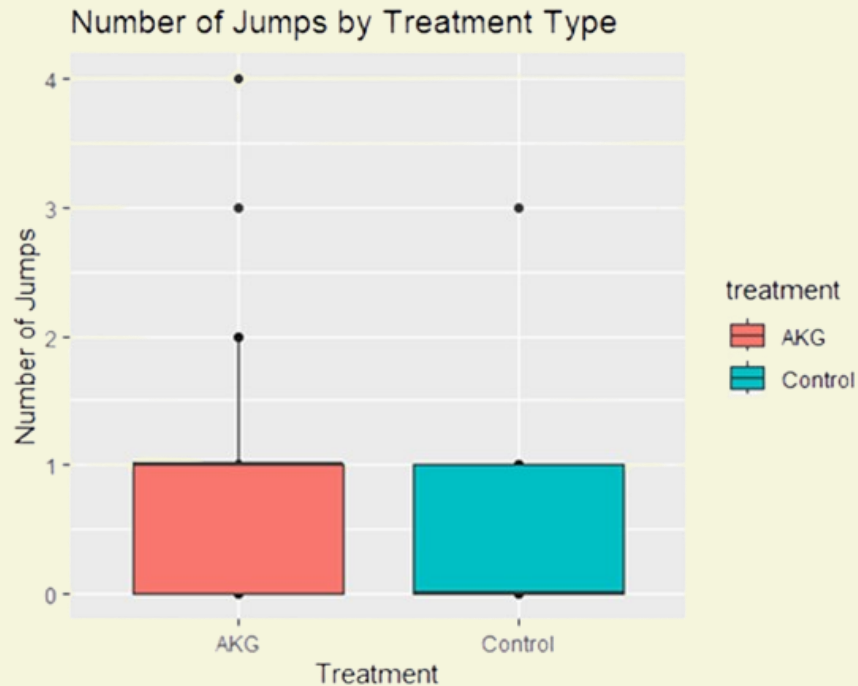
```
library(ggplot2)
```

# The Results

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

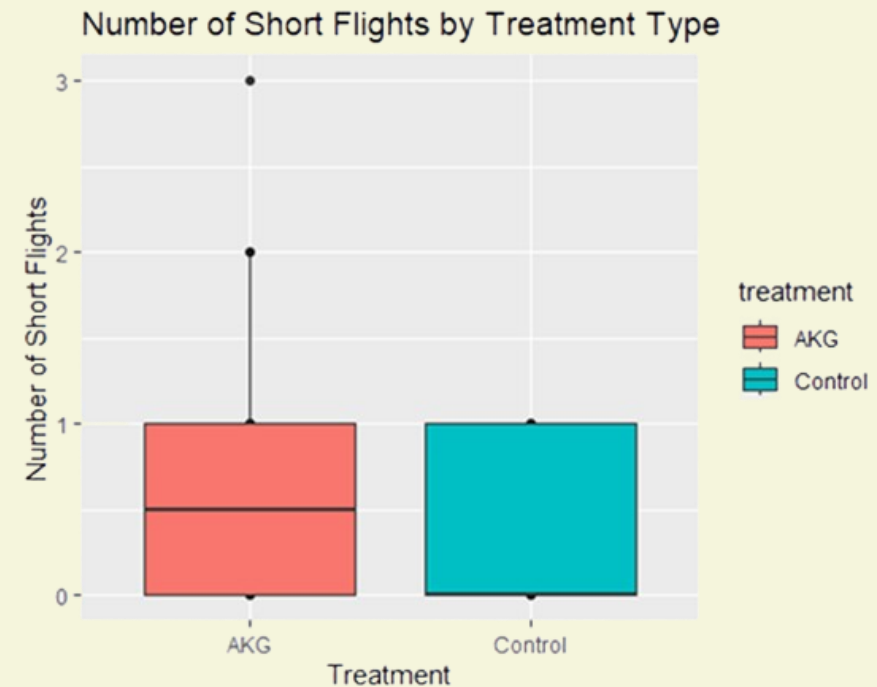
## # Boxplot for the number of jumps

```
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")
```



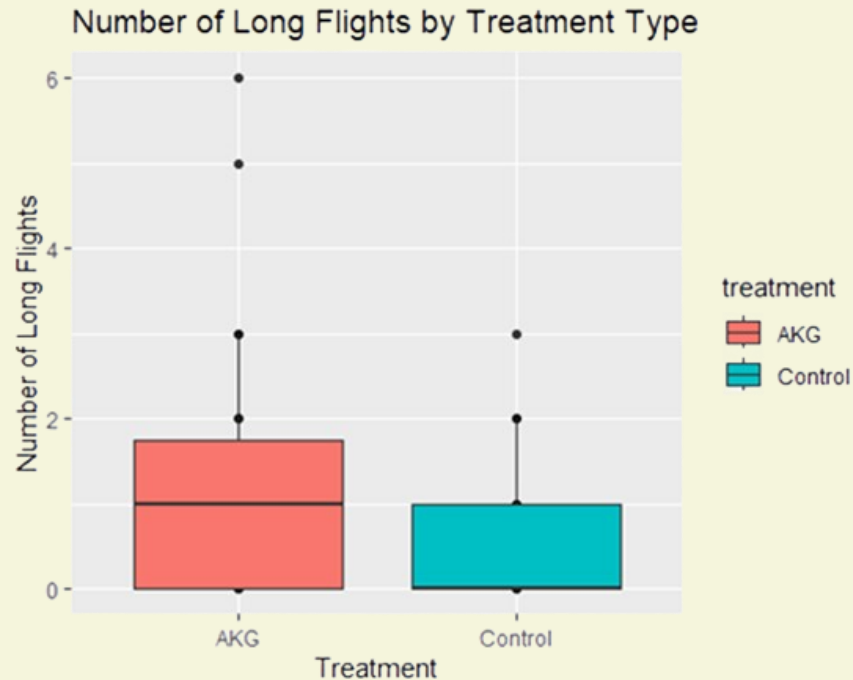


# The Results

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

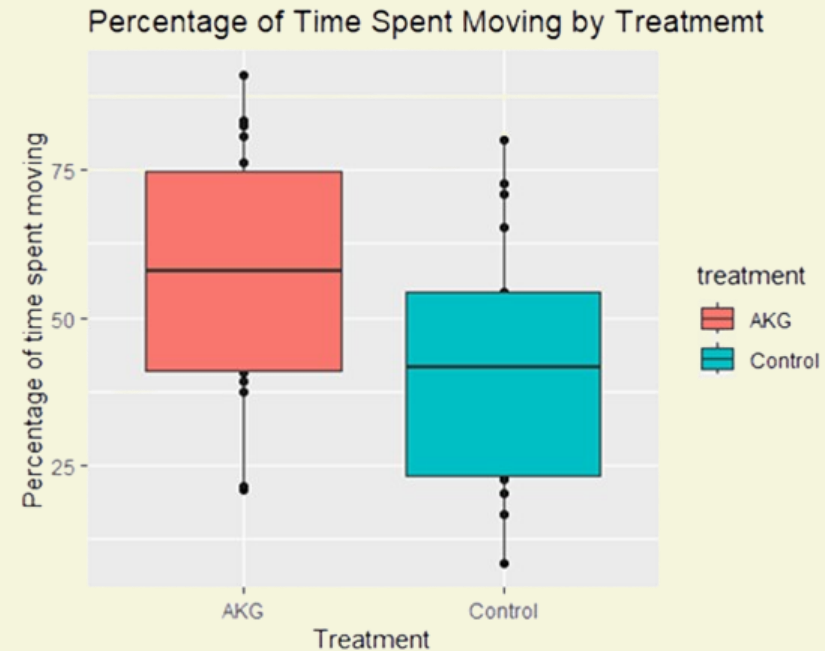
## # Boxplot for the number of long flights

```
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"))

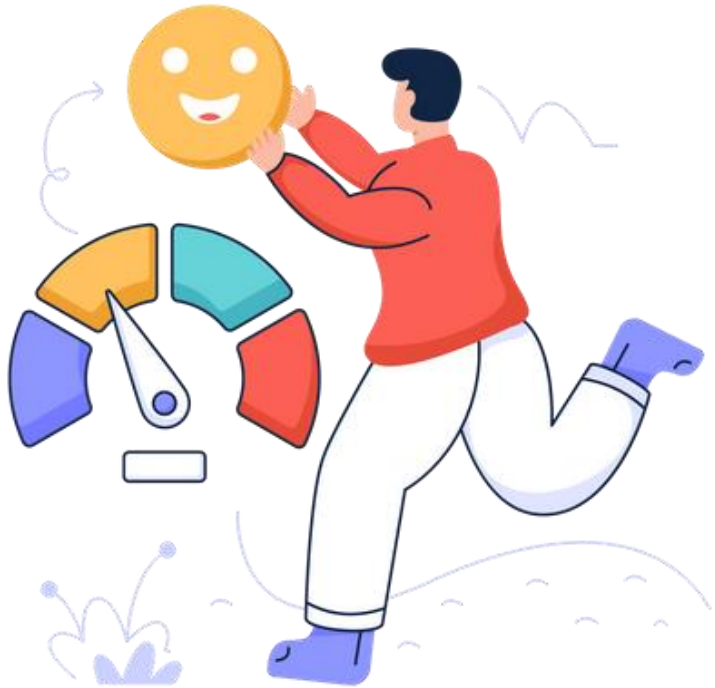
## Warning in dpois(y, mu, log = TRUE): non-integer x = 37.500000
## Warning in dpois(y, mu, log = TRUE): non-integer x = 74.333333
## Warning in dpois(y, mu, log = TRUE): non-integer x = 72.166667
## Warning in dpois(y, mu, log = TRUE): non-integer x = 44.762063
## Warning in dpois(y, mu, log = TRUE): 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:

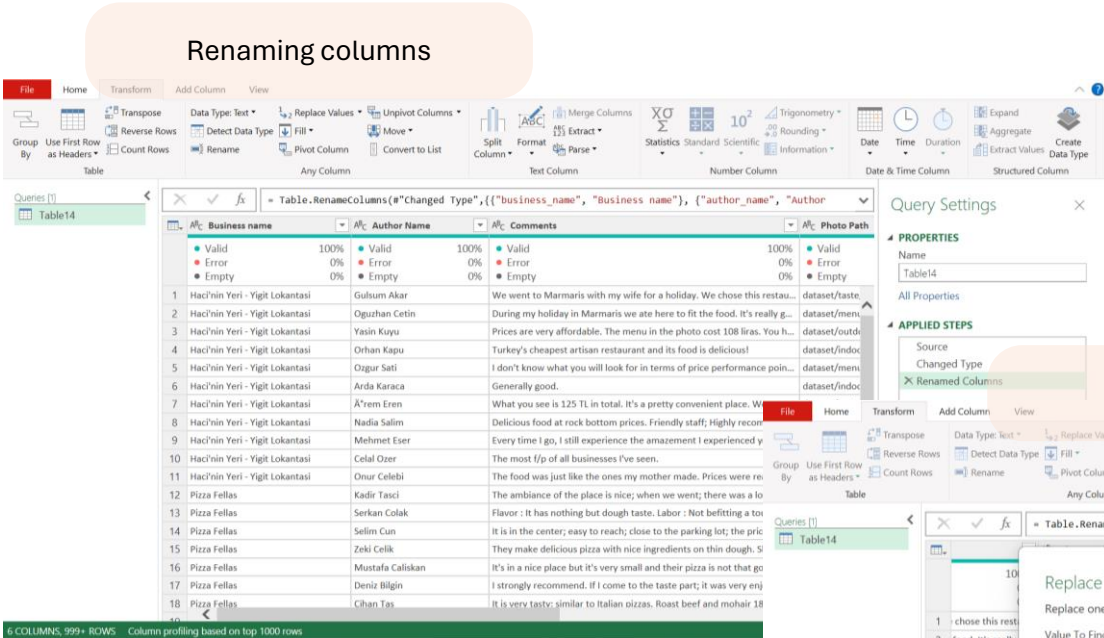
- Restaurant names
- Author names
  - Text
  - Photos
  - Rating
  - Rating category - considers *taste, menu, and indoor and outdoor atmosphere*.

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

# Cleaning Process

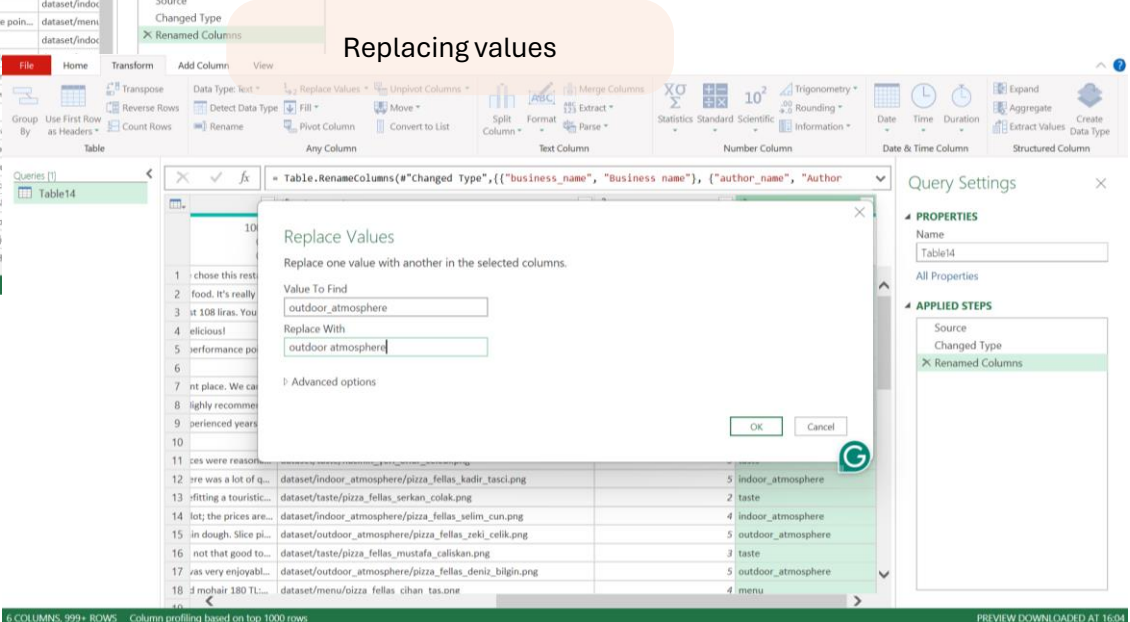
## The Customer Satisfaction Project

### Renaming columns



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

### Replacing values



Value To Find	Replace With
outdoor_atmosphere	outdoor atmosphere

Advanced options

Value To Find	Replace With
dataset/indoor_atmosphere/pizza_fellas_kadir_tasci.png	5 indoor_atmosphere
dataset/taste/pizza_fellas_serkan_colak.png	2 taste
dataset/indoor_atmosphere/pizza_fellas_selim_cun.png	4 indoor_atmosphere
dataset/outdoor_atmosphere/pizza_fellas_zeki_celik.png	5 outdoor_atmosphere
dataset/taste/pizza_fellas_mustafa_caliskan.png	3 taste
dataset/outdoor_atmosphere/pizza_fellas_denis_bilgin.png	5 outdoor_atmosphere
dataset/menu/pizza_fellas_cihan_tas.one	4 menu

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

e.g.

outdoor\_atmosphere →  
outdoor atmosphere



# Cleaning Process

## The Customer Satisfaction Project

Adding conditional a column

Column Name	Operator	Value	Output
Rating Score	is less than or equ...	2	Negative
Rating Score	equals	3	Okay
Rating Score	is greater than or ...	4	Positive

Removing a column

Query Settings

PROPERTIES

Name: Table14

APPLIED STEPS

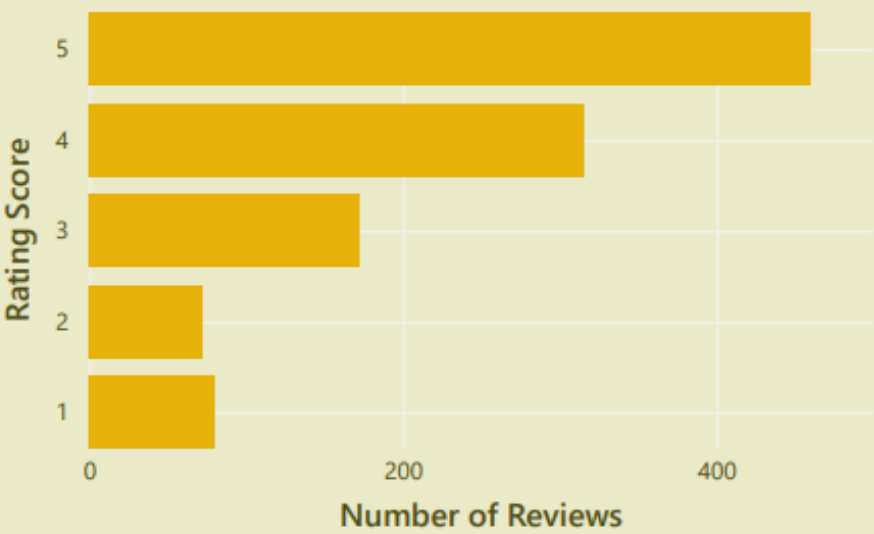
- Source
- Changed Type
- Renamed Columns
- Replaced Value
- Replaced Value1
- Added Conditional Column

- 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

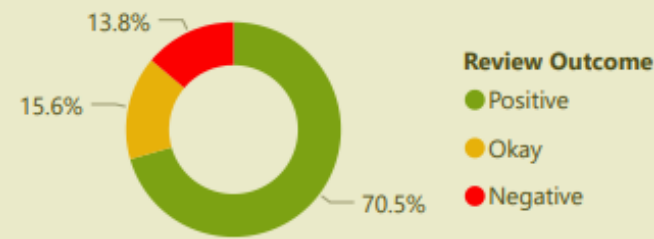


Restaurants

Abidin Tantuni

Ahsap Iskender

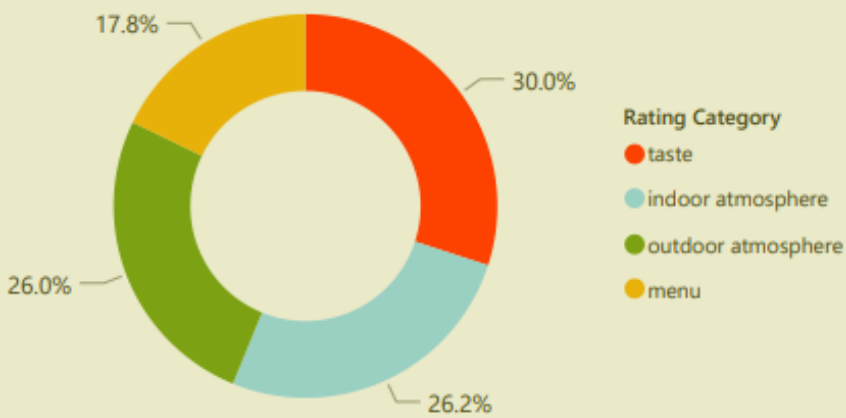
Customer Satisfaction



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

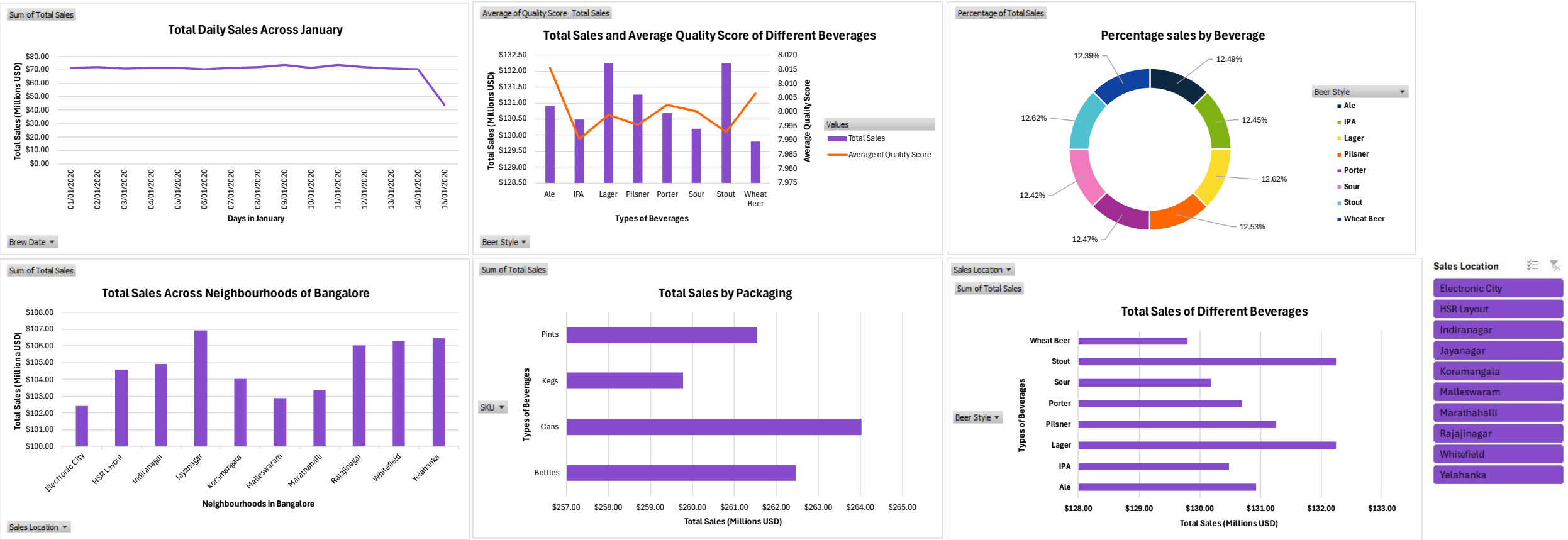
Kegs  
\$259.8 million





# Sales Breakdown

## The Brewery Operations Project



# Operations Breakdown

## The Brewery Operations Project



**SKU With The Highest Quality Score**  
Bottles  
8.0099/10.0000



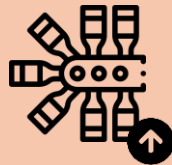
**Beverage With The Lowest Quality Score**  
IPA  
7.9905/10.0000  
\*Stout – close second



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



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



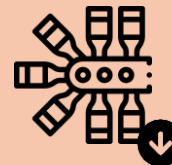
**Most produced SKU**  
Cans  
69.4 million litres  
\*Kegs – close second



**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



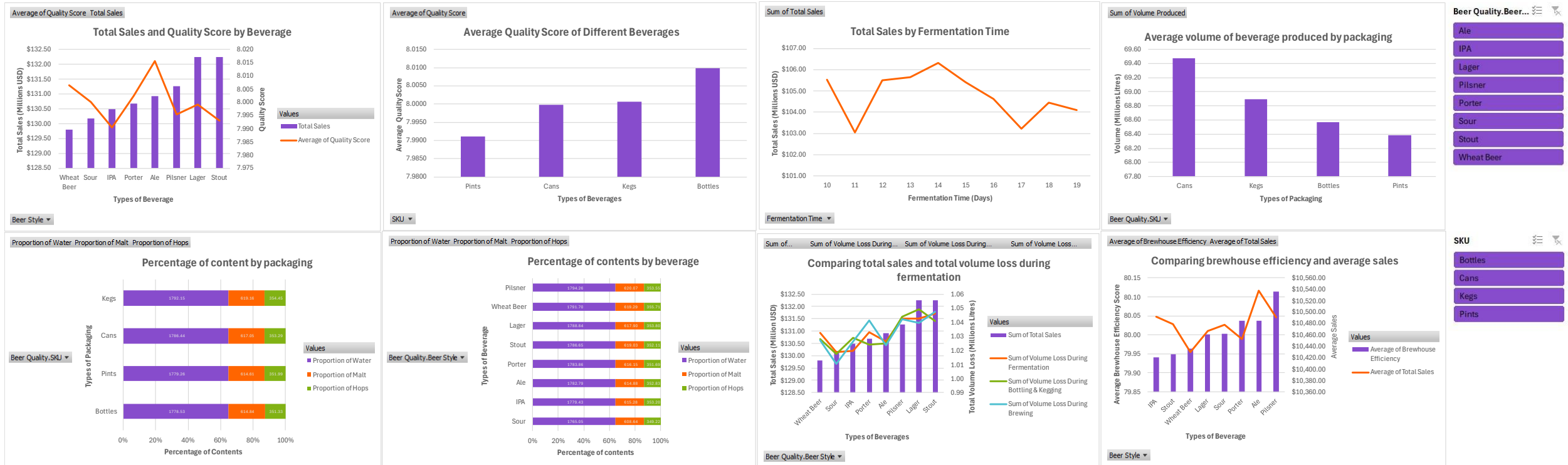
**Least produced SKU**  
Pints  
68.3 million litres



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

# Operations Breakdown

## The Brewery Operations Project



### 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 Workbook Links

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 Advanced Editor

Close & Load Refresh Preview Manage Query

Choose Columns Remove Columns Manage Columns Reduce Rows Sort Split Column Group By Replace Values Transform

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

023

2

3

OK Cancel

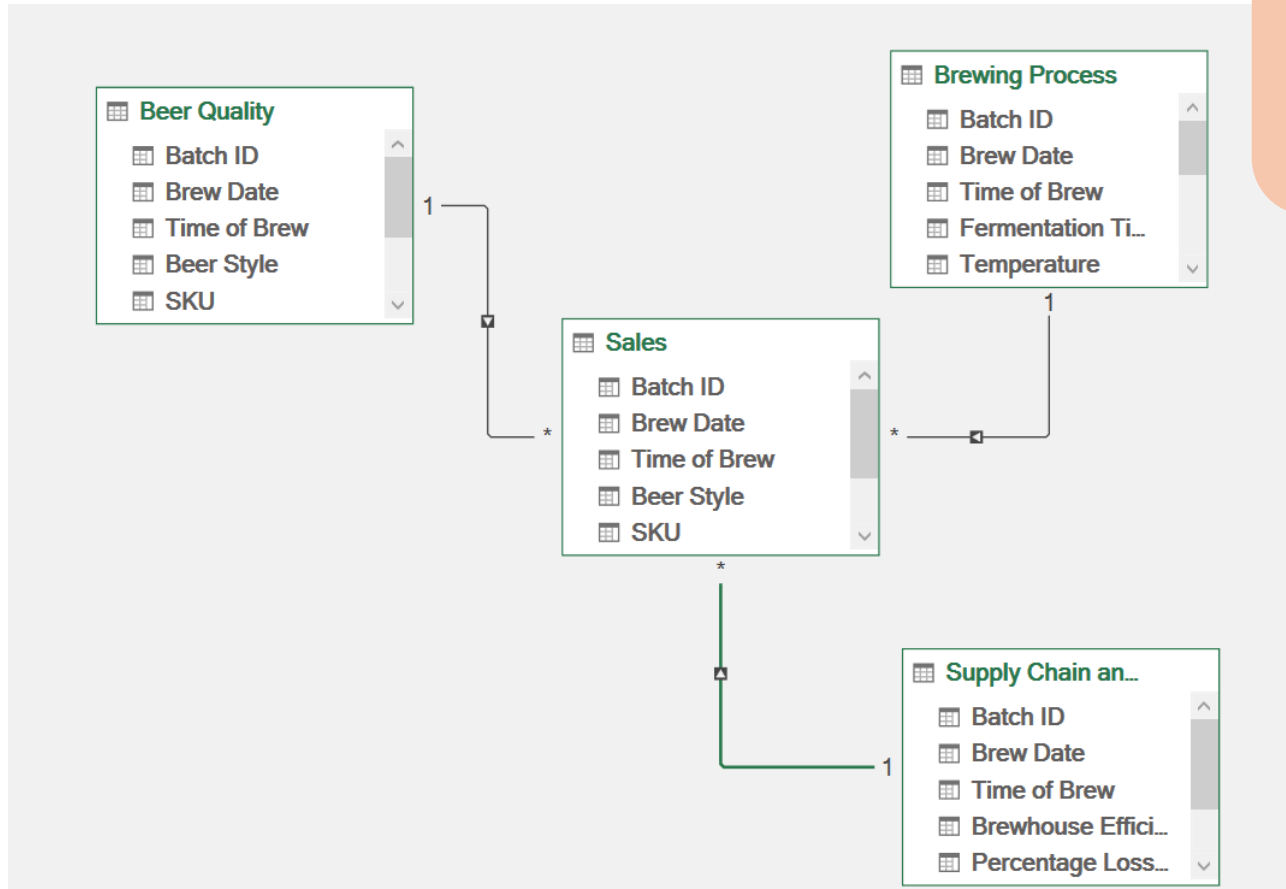
Yeanie Lamin



# The Data Model

## The Brewery Operations Project

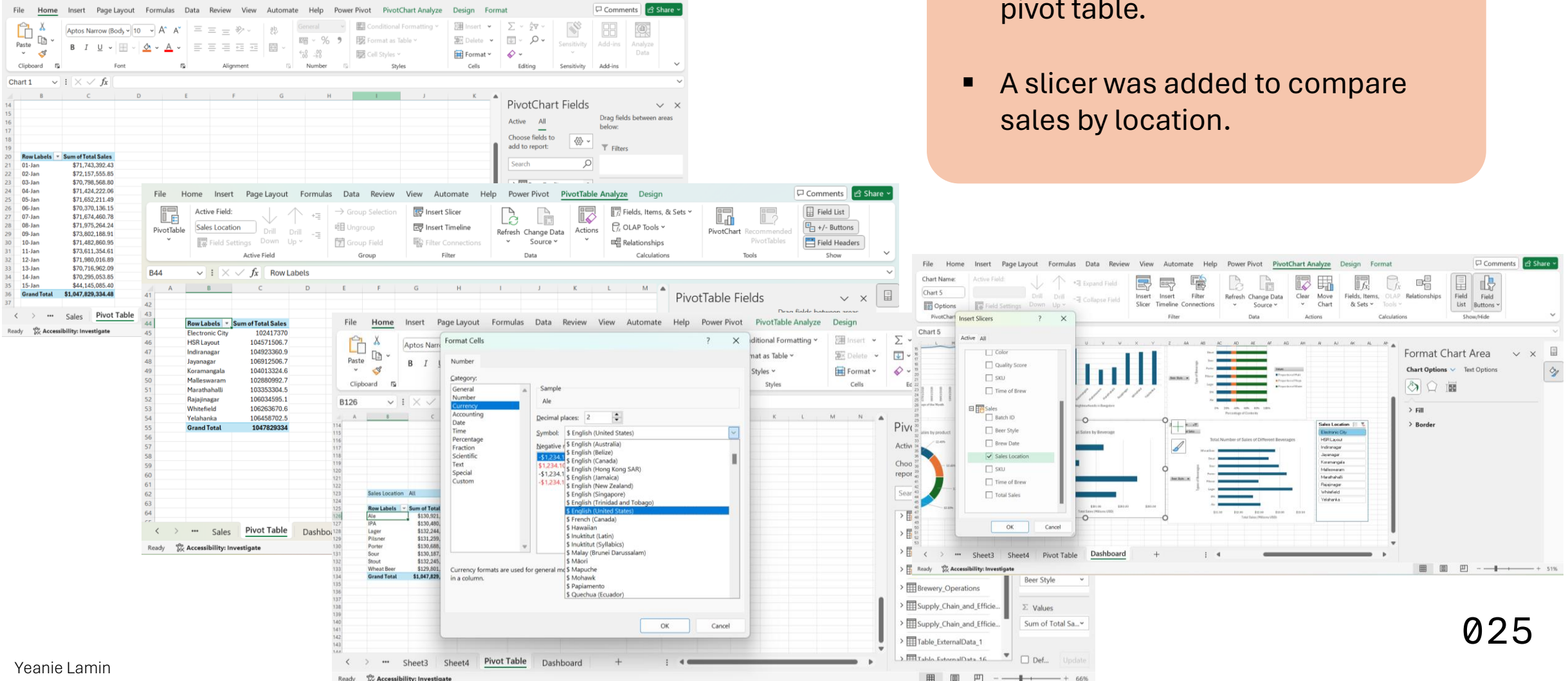
- Due to the large number of columns in the dataset, I broke the dataset down into 4 smaller datasets.
- I then created a data model.



# 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!

