



INDIAN INSTITUTE
OF TECHNOLOGY
PALAKKAD

Applied Data Science

Capstone Project Proposal– Retail

Capstone Project Background and Objective



Plato's Pizza is a Greek-inspired pizza place in New Jersey. The restaurant has been collecting transactional data for the past year, but haven't been able to put it to good use. They would like to find out opportunities to drive more sales and work more efficiently.

As a **Data Consultant** hired by the restaurant to find useful insights to improve operations – these are some queries to answer – best and worst selling pizza, average order value, busy days / times, current seat utilization etc.

ML models to forecast pizza sales revenue and pizza counts sold are built in order to help the restaurant with personnel planning and improving sales.



Capstone Project: Approach

Approach Steps:

1. Check the dataset: available as 4 csv files of entire restaurant transactional database of 4 tables.
2. Load to SQL tables for ease of joining and removing unnecessary columns. Export as a combined csv file for processing using Python
3. Use of summary statistics, EDA and visualizations to answer some of the basic queries posed by business.
4. Use of time series model to analyse the sales/revenue pattern of the restaurant
5. Use of regression model to determine an equation to forecast the no. of pizzas to be made daily based on the results of step 4.
6. Provide analysis and recommendations based on the above steps.

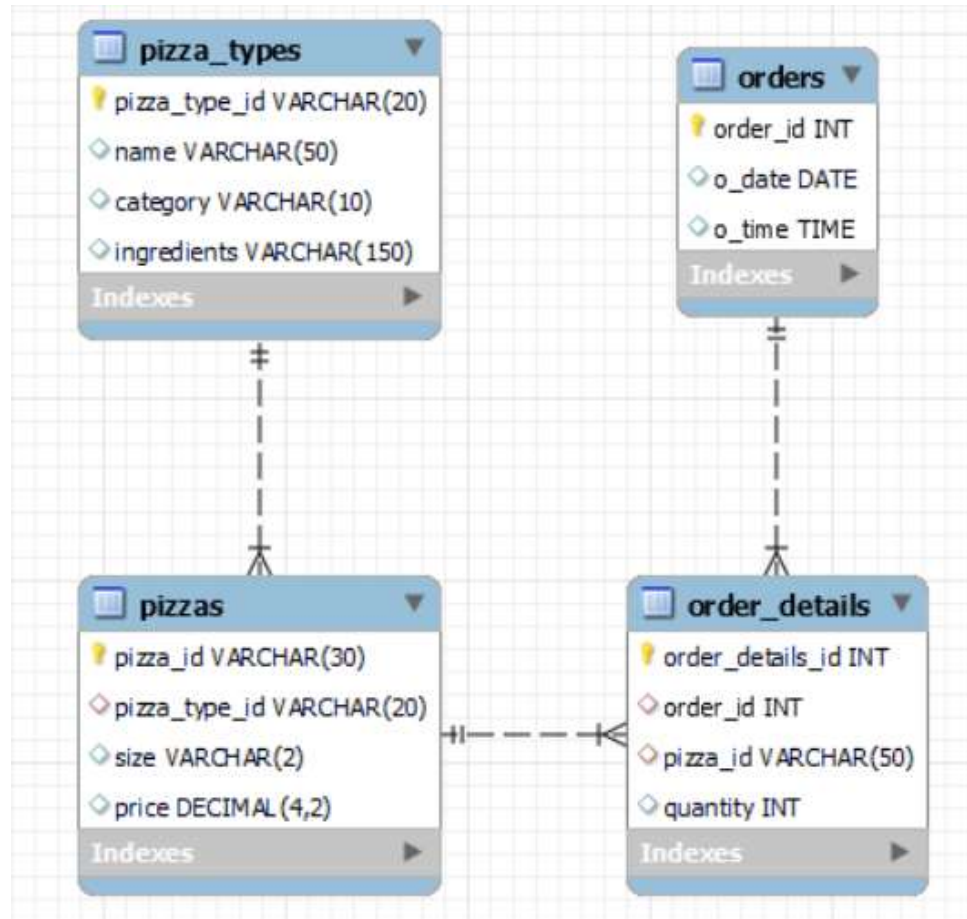
Resource Requirements:

- Tools: SQL, Python and related ML, statistics and visualization packages
- Data : Pizza Place Sales [dataset](#)



Dataset and Pre-processing Details

The dataset available in the form of 4 csv files were loaded in to local MySQL database to get the below tables for processing and necessary information were extracted as a single csv file for further processing using Python.



EER Diagram

Table	Field	Description
orders	order_id	Unique identifier for each order placed by a table
	o_date	Date the order was placed
	o_time	Time the order was placed
order_details	order_details_id	Unique identifier for each pizza placed within each order (pizzas of the same type and size are kept in the same row, and the quantity increases)
	order_id	Foreign key that ties the details in each order to the order itself
	pizza_id	Foreign key that ties the pizza ordered to its details, like size and price
	quantity	Quantity ordered for each pizza of the same type and size
pizzas	pizza_id	Unique identifier for each pizza (constituted by its type and size)
	pizza_type_id	Foreign key that ties each pizza to its broader pizza type
	size	Size of the pizza (Small, Medium, Large, X Large, or XX Large)
	price	Price of the pizza in USD
pizza_types	pizza_type_id	Unique identifier for each pizza type
	name	Name of the pizza as shown in the menu
	category	Category that the pizza fall under in the menu (Classic, Chicken, Supreme, or Veggie)
	ingredients	Comma-delimited ingredients used in the pizza as shown in the menu (they all include Mozzarella Cheese, even if not specified; and they all include Tomato Sauce, unless another sauce is specified)

Data Dictionary



Order Summary Analysis



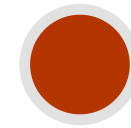
Avg. Order Value
38.27\$



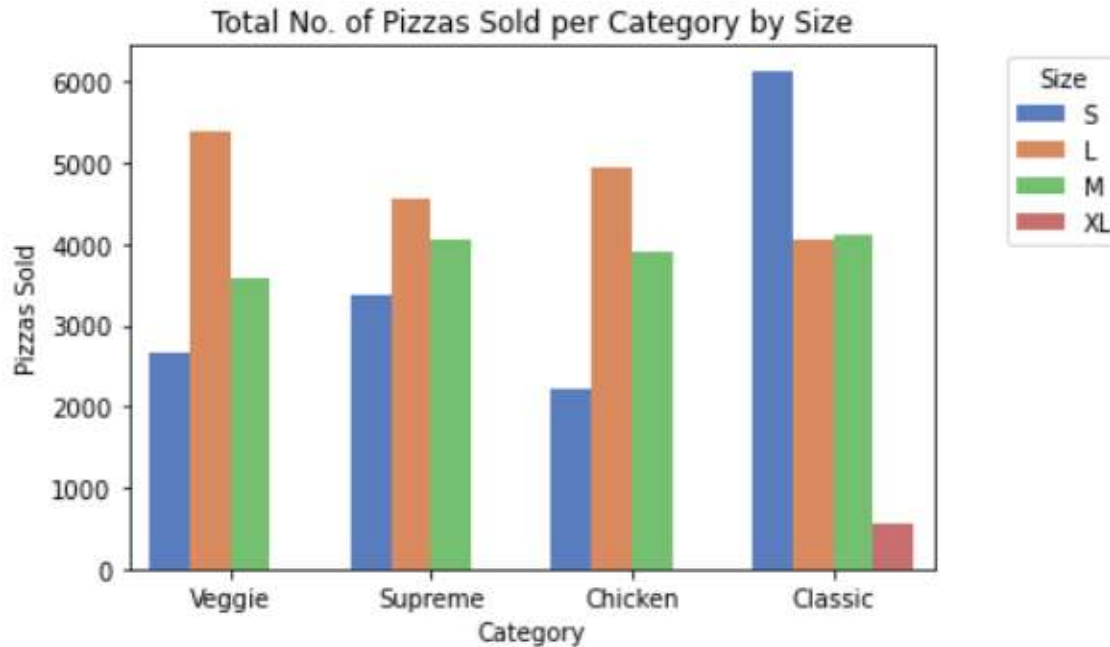
Total Orders
21350



Total Pizzas Made
49546



Total Revenue
816853.45\$



Observations:

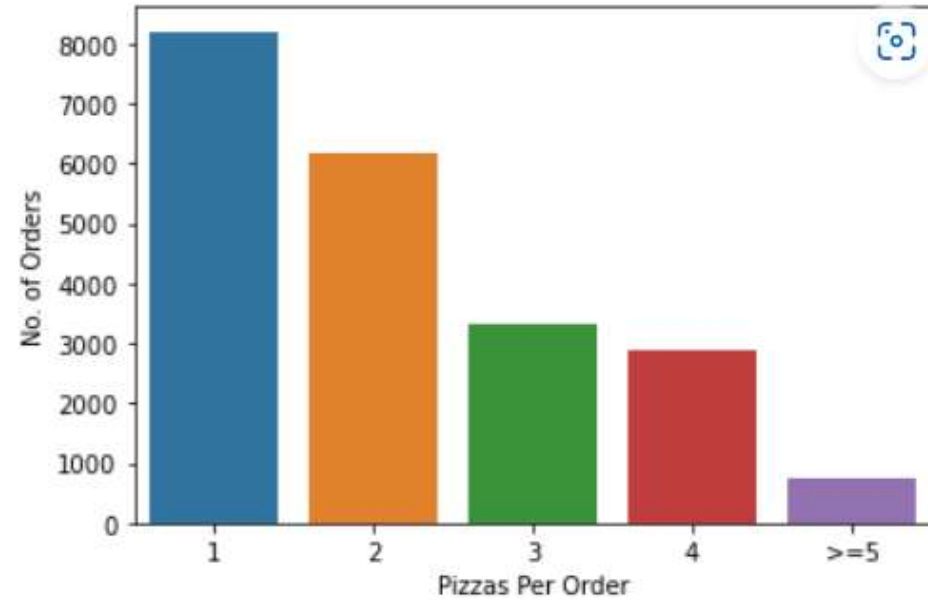
- Transactional data is available for the year 2015
- Customers generally prefer to order 'S' or 'Small' size of 'Classic' pizzas when compared to other sizes and categories.
- 'XL' size is offered only in one pizza in the 'Classic' category – 'The Greek Pizza'



Break-Up of Total #Pizzas Made

Pizza Size	Counts
L	18956
M	15635
S	14403
XL	552

Pizza Category	Counts
Classic	14860
Supreme	11987
Veggie	11649
Chicken	11050



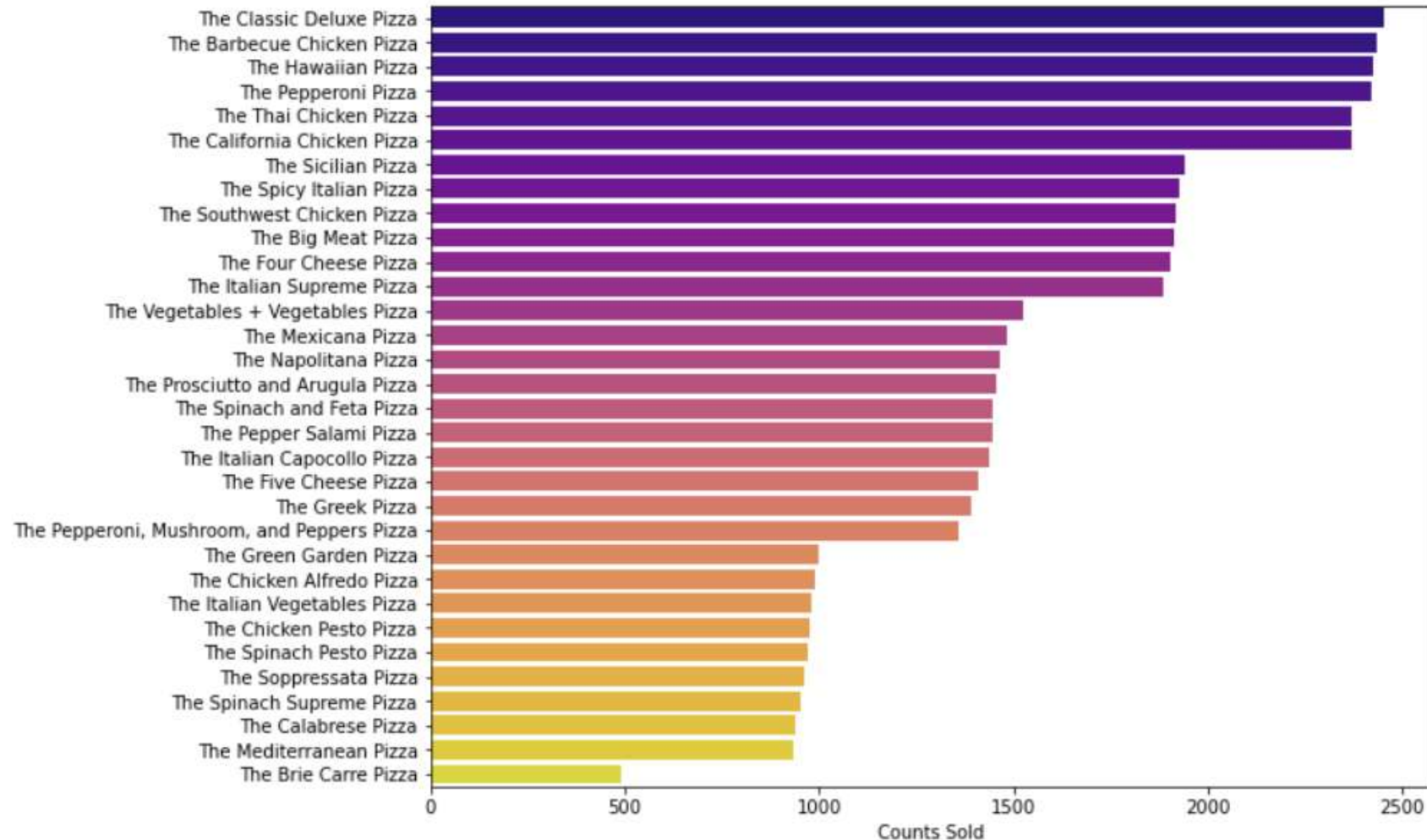
Observations:

- 552 'XL' size pizzas of 'The Greek Pizza' were sold in the year.
- Total of 14860 'Classic' pizzas were sold in the year.
- The most ordered size of pizza was 'L' with 18956 counts.
- Majority of the orders are for 1 or 2 pizzas.



More Pizza Stats : Counts Sold

Plato's Pizza sells 32 different pizzas in different sizes.



Best Seller:

'The Classic Deluxe Pizza'

Worst Seller:

'The Brie Carre Pizza'



More Pizza Stats : Revenue



The Greek Pizza	Size - XL	25.5
The Brie Carre Pizza	Size - S	23.65



The Pepperoni Pizza	Size - S	9.75
The Hawaiian Pizza	Size - S	10.5



Highest: The Thai Chicken Pizza - \$43434.25

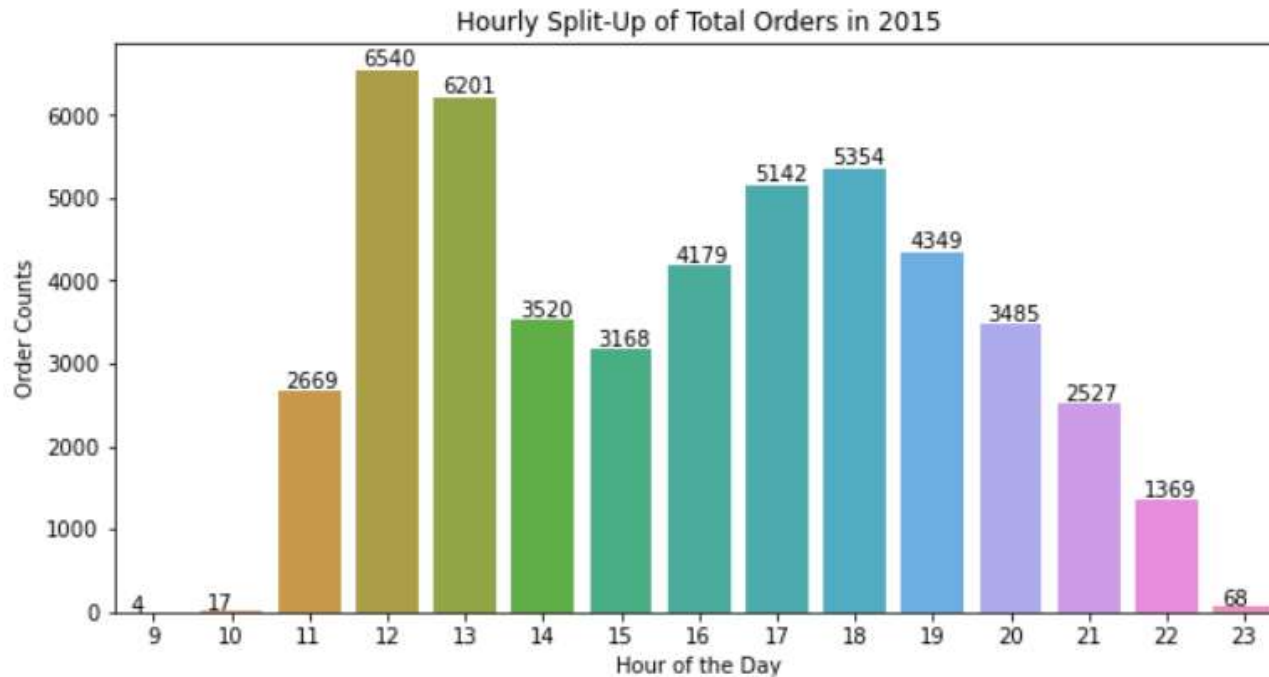
Lowest: The Brie Carre Pizza - \$11588.5

Observations:

- 'The Brie Carre Pizza' is available only in size 'S' and is one of the most expensive pizzas sold. This seems to be a reason for lower counts being sold.
- It can be seen that some of the lower priced pizzas feature in the top selling pizzas. Even though the pizza contributing highest to revenue is not the best seller pizza, it is in the top-5 highest selling pizzas.



Busy Times of the Day



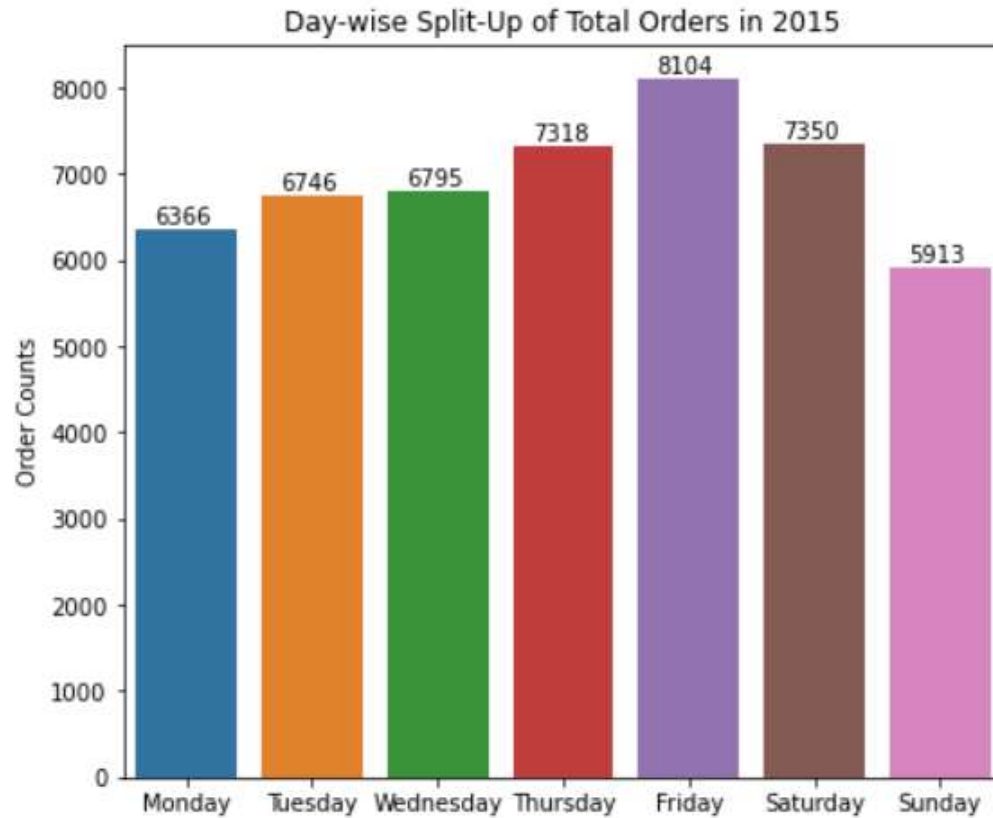
Busiest Hours of the Day:

Lunch Hours: 12:00 Noon to 2:00 PM

Dinner Hours: 5:00 PM to 7:00 PM

Observation: Most of the days – orders peak around lunch and dinner.

Busy Days of The Week



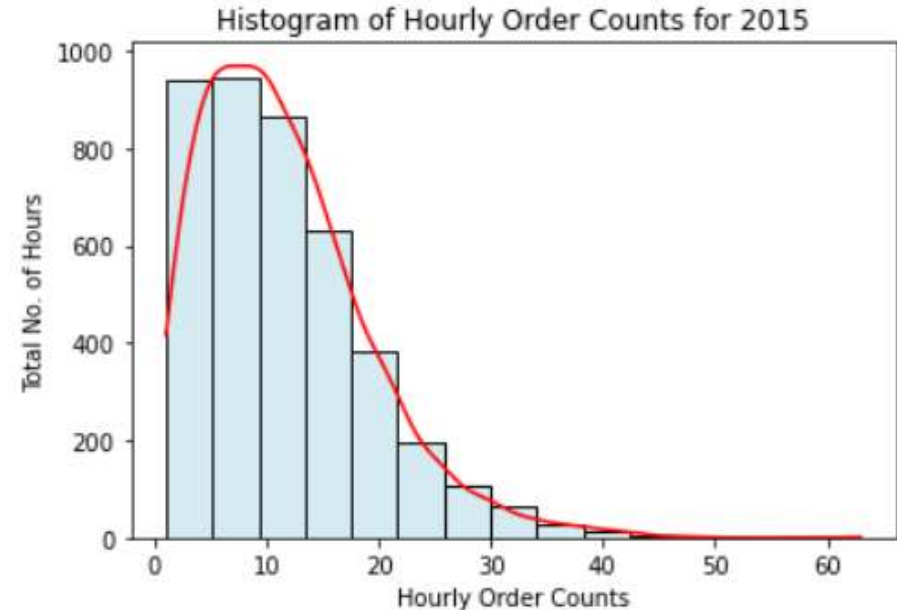
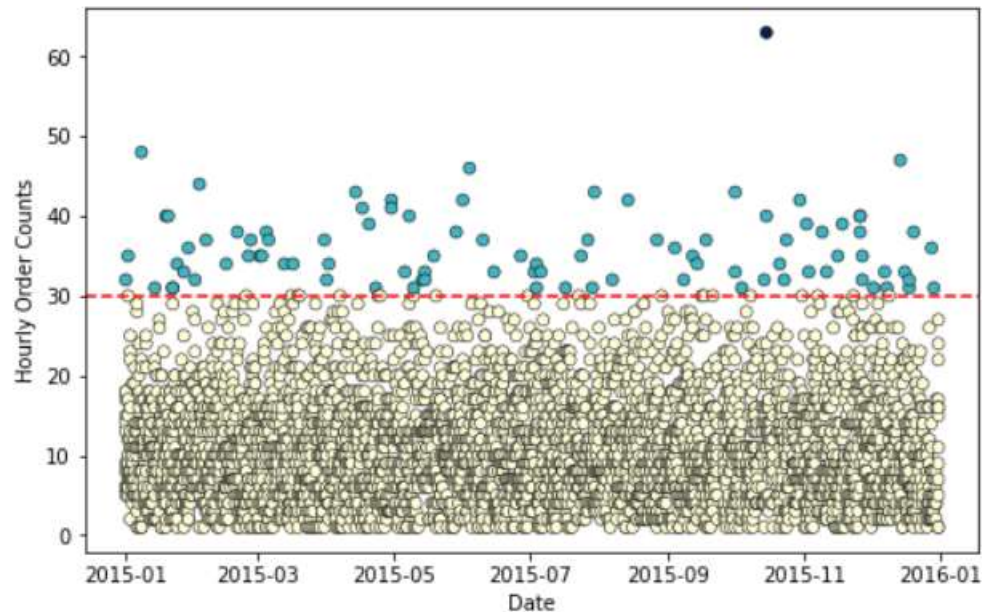
Busiest Days of the Week:

Friday, Saturday and Thursday



Seat Utilization – Part 1

Plato's pizza has 15 tables with 4 seats (max: 60 people).

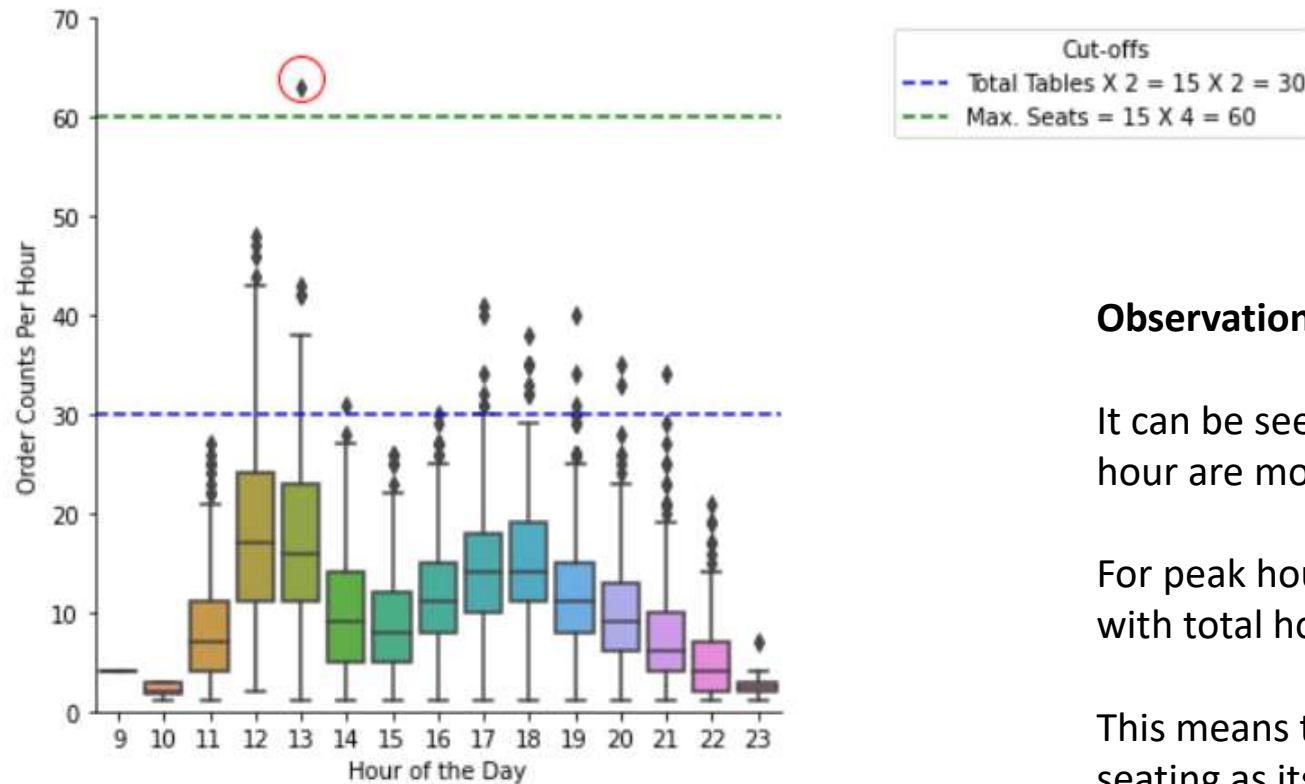


- Assuming that on an average people spend a maximum of 25-30 minutes in the pizza shop. This means that in an hour, #30 orders can be easily accommodated [it is given that order id is specific to tables].
- When the orders occasionally peak above this count, it looks like each of 60 seats have been assigned to different order ids as needed.
- In the entire year, there were 84 days with 94 hours which had total hourly order counts more than 30.



Seat Utilization – Part 2

This plot gives an idea of the number of total orders that have been received per hour across all the days.



Observation:

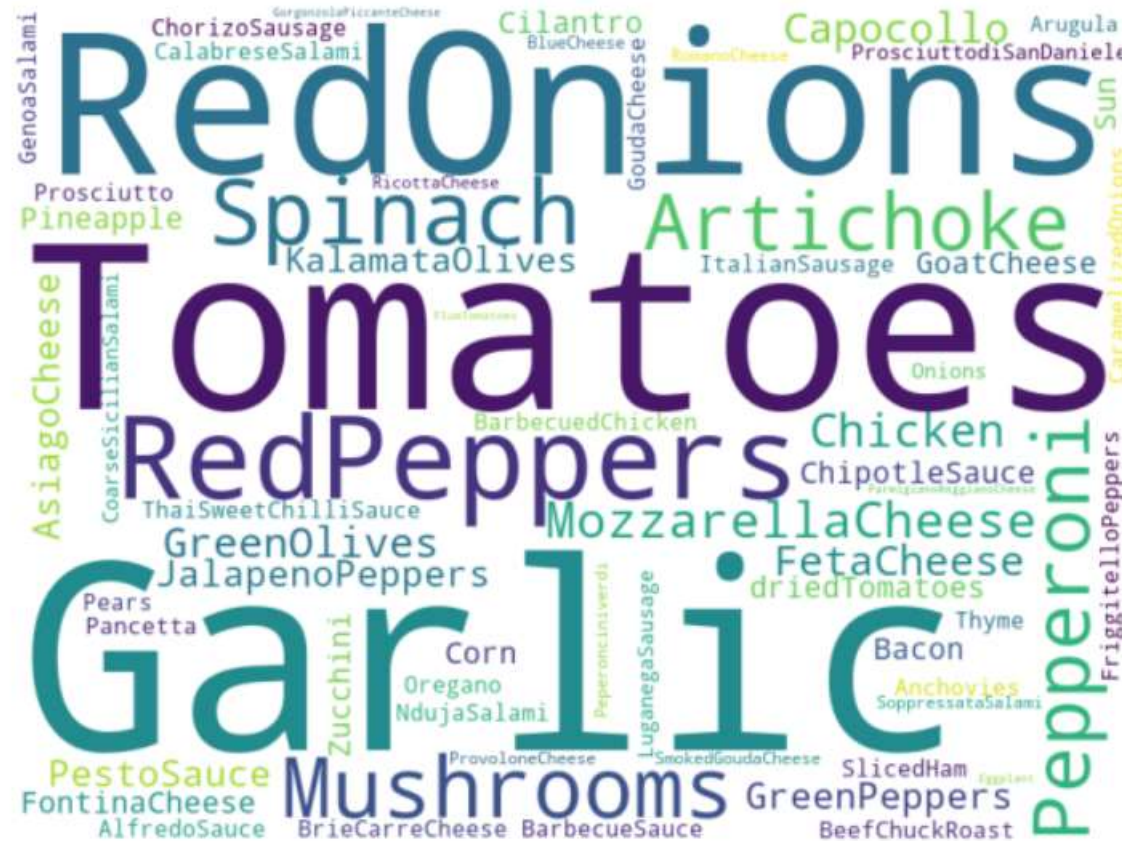
It can be seen that, except for lunch hours, no. of total orders per hour are mostly within 30.

For peak hour orders, in the entire year, there is only one instance with total hourly order above 60 (the total seat count).

This means that the restaurant needn't worry about re-sizing the seating as its adequate for the orders.



Pizza Ingredients and Toppings Analysis



Observation:

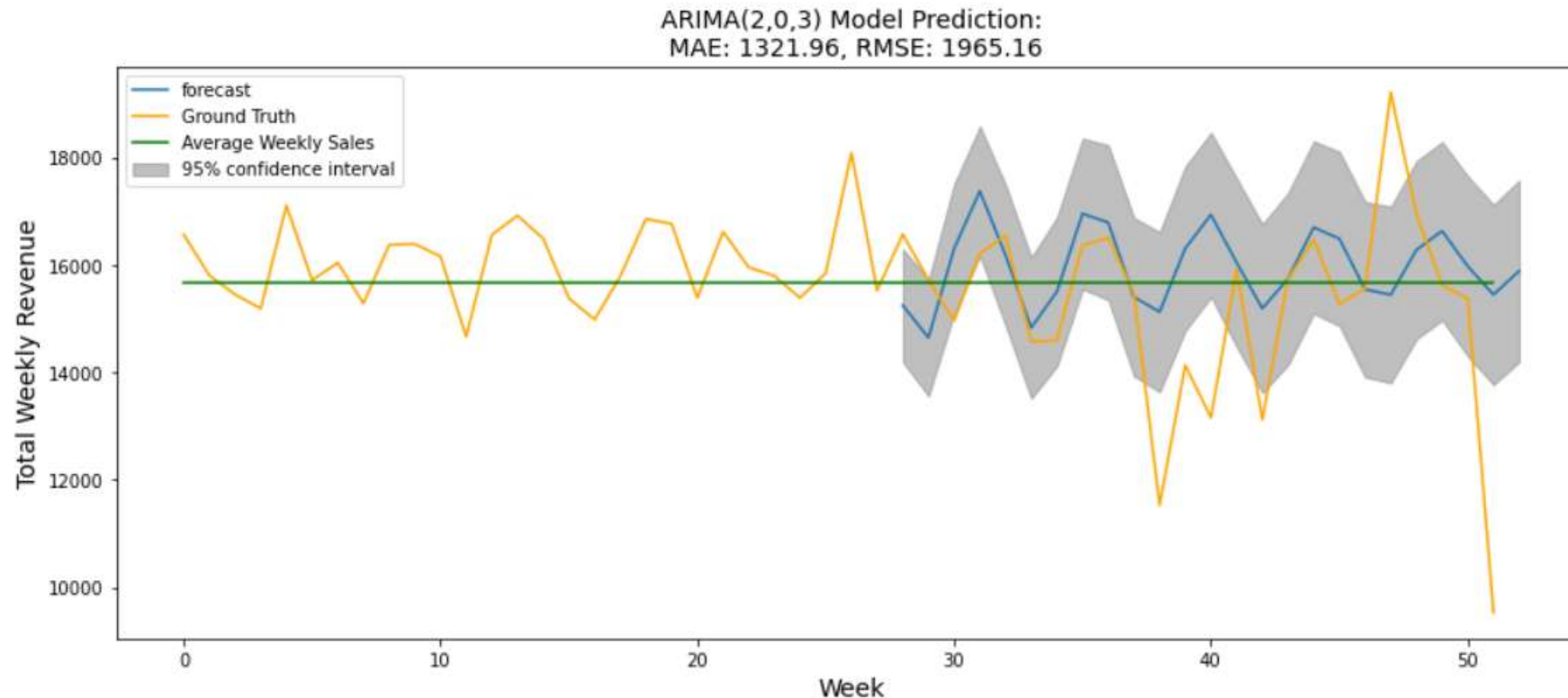
Of the listed ingredients – most pizzas contain Tomatoes, Garlic, Red Onions along with - Mozzarella Cheese and Tomato Sauce. Interestingly – Spinach is also a popular ingredient!

ML Models for Weekly Revenue Analysis and Estimating Pizza Counts

Model Pre-Processing and Further Steps:

- Weekly revenue analysis and forecast is done using **ARIMA model**
- Data is in the form of transactions – one entry per order and one entry for each item in the order.
- This is aggregated to daily sales to get – Date, Order Counts, Pizza Counts and Total Revenue
- This data is stationary. It is then downsampled to week-wise data to perform weekly revenue analysis
- The shop was closed on 7 days in the year:
 - '2015-09-24' (Thursday), '2015-09-25' (Friday) – in September
 - '2015-10-05' (Monday), '2015-10-12' (Monday), '2015-10-19' (Monday), '2015-10-26' (Monday) – in October
 - '2015-12-25' (Friday) - in December
- The business closure days sales were not interpolated / imputed so as to capture zero sales days in the model.
- Model parameters ARIMA (p,d,q) were estimated as p=2, q=3, d=0 using PACF/ACF plots and a few combinations were tried to arrive at final parameters by checking and comparing the r2_score and RMSE values of the combinations.
- The relationship between no. of pizzas sold daily and daily revenue is fit using a **linear regression model**. This is necessary as the **pizzas are not uniformly priced**.
- The **contribution of each day to the week's sales** is used along with the **parameters from the linear model** to estimate the **no. of pizzas** to be made.

Results : Weekly Revenue Analysis – Time Series Model

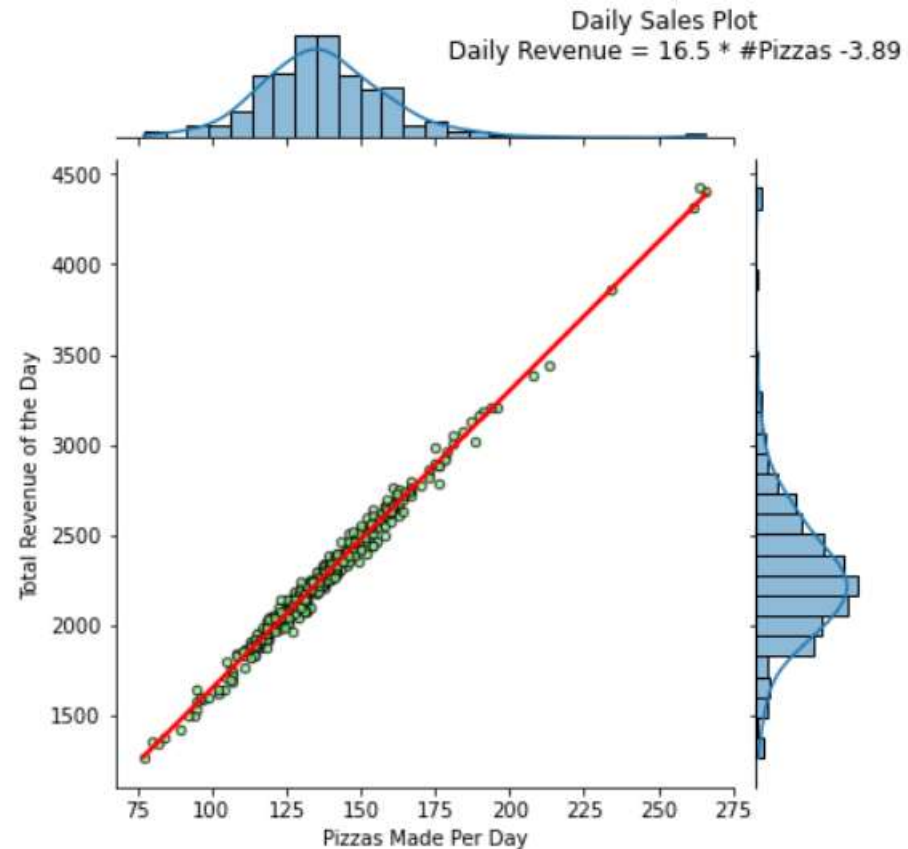


Observation:

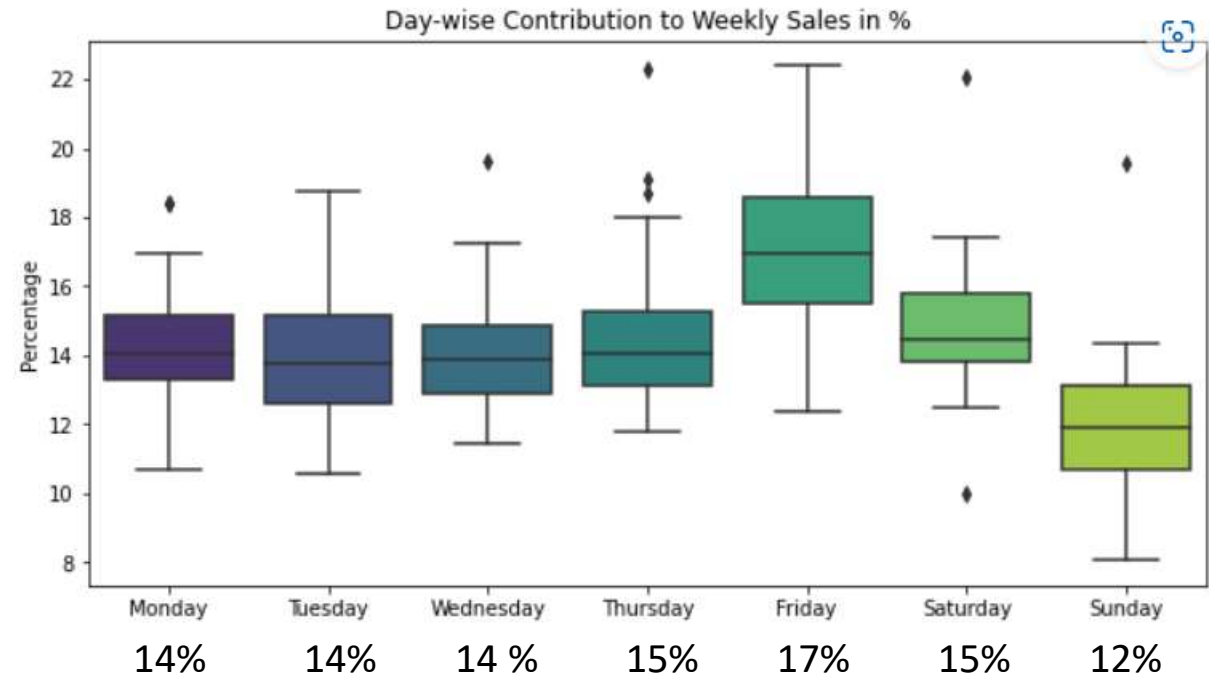
As the data is only of a single year – the model is able to more or less predict the weekly revenue accurately compared to the average weekly sales. There are fluctuations in prediction due to the following reasons – 1) Shop was closed on all Mondays in October 2) Shop closure on 2 consecutive days in a week in September 3) 53rd week of the year contains only 31st Dec, 2015.



Results: Estimate #. of Pizzas Sold Per Day of Week



It can be seen that the no. of pizzas made on a day and the day's total revenue are linearly related (equation on figure title)



Friday sales contributes the maximum to weekly revenue.
Sunday sales contributes the minimum to weekly revenue.

ML Model Summary

Forecasted weekly revenue values can be split based on the day's % contribution and used to predict the no. of pizzas to be made on a day in the week :

Daily Revenue Forecast = Weekly Revenue Forecast X Day's Contribution %

#No. of Pizzas To Be Made = $\frac{\text{Daily Revenue Forecast} + 3.89}{16.5}$

Challenges:

As the data is available only for one year the essence of seasonality and trend could not be very well seen. As we collect more data, the model can be further refined to do better prediction.



Recommendations - 1

Opening Hours:

- Current opening hours are 9am - 12:00 midnight.
- From Slide 9: in the entire year – there has been only a total of $4+17+68 = 89$ orders during 9am-10am, 10am-11am, 11pm-12:00 midnight slots. Suggestion to **change opening hours** to **11am -11pm** for better efficiency and reducing utilities costs

Seat Utilization:

- Looking at visualizations from Slides 6,11 and 12 – the current seat availability is good for most of the hours of the day.
- Seat utilization can be further improved by converting a few tables to **counter top (single customer) service** and adding **2-seat tables** to accommodate the analysis that most orders are predominantly of 1 or 2 pizzas. This can reduce empty seats and help in catering to more orders per hour.

Worst Selling Pizza:

- From Slides 7,8: Provide **promotional offers** to improve sales of 'The Brie Carre Pizza' and market it as Gourmet pizza.

Staffing:

- From Slides 9, 10: Plan for more staff to handle **peak hours during lunch and dinner** – 12:00 Noon -2:00 PM and 5:00 PM – 7:00 PM and for **busy days** – Thursday, Friday and Saturday.



Recommendations - 2

Non-Working Days

- Consecutive business closed days affect the weekly revenue significantly as seen from Slide 16
- Keep maintenance works to **Sundays**, as it is the lowest contributor to weekly revenue. Refer Slide 17
- From the sales pattern – the pizza place location is possibly in a commercial area. The non-working days may coincide with general business holidays to reduce loss of revenue.

No. of Pizzas to Make

- The equation from **Slide 18** can help in planning the no. of pizzas to be made.
- From Slide 17: A minimum of 124 pizzas are made on most of the days.





INDIAN INSTITUTE
OF TECHNOLOGY
PALAKKAD