

Mini Project Report

เรื่อง (Subject) : Pizza Sales Insights, What the Data Tells Us

แหล่งที่มาของข้อมูล (Data Source) : Maven Analytics website :

https://www.mavenanalytics.io/dataplayground?order=date_added%2Cdesc&search=pizza

รายละเอียดชุดข้อมูล (Dataset) : เป็นข้อมูลเกี่ยวกับการขายของร้าน pizza แหล่งหนึ่งซึ่งมีข้อมูลเพียง 1 ปี โดยจะมีข้อมูลวันและเวลาในการสั่งซื้อแต่ละครั้ง พร้อมรายละเอียดเพิ่มเติมเกี่ยวกับประเภท ขนาด ปริมาณ ราคา และส่วนผสม

วิธีการดำเนินการ แบ่งการดำเนินการเป็น 2 วิธีการ คือ

1. การจัดการและทำความสะอาดข้อมูลกับเพื่อให้ได้ data frame ที่ต้องการ
2. การทำ Visualization เพื่อหาคำตอบหรือ Insight ตามสมมติฐานที่คาดการณ์ไว้

วิธีการดำเนินการที่ 1 : จัดการและทำความสะอาดข้อมูลกับเพื่อให้ได้ data frame ที่ต้องการ

ซึ่งนำเนินการตามไฟล์งานที่ 1 : “ **CleaningCode_PizzaProject** ” โดยมีหลักการโดยสรุปดังนี้

เชื่อมข้อมูลเข้าหากัน เนื่องจากเรามี dataset ย่อยที่ได้จากแหล่งข้อมูล ดังนี้ order_details.csv, orders.csv, pizza_types.csv และ pizzas.csv (โดยรายละเอียดของข้อมูลสามารถดูได้จากไฟล์ Meta Data.) อ่าน data frame ของ dataset ย่อยทั้ง 4 dataset เพื่อตรวจสอบข้อมูลเบื้องต้น เช่น ตรวจสอบ missing value, หา primary key และ foreign key สำหรับกระบวนการเชื่อมข้อมูลต่อไป จากการสำรวจพบว่า : ทุก data frame ไม่มีค่า missing value และทุก dataset มี primary key และ foreign key ที่เชื่อมข้อมูลเข้าด้วยกันได้จึงเชื่อมข้อมูลเข้าเป็น data frame เดียว

```
1 pizza_1 = order_details.merge(orders, left_on='order_id',right_on='order_id')
2 # pizza_1.head(3)
3 pizza_2 = pizza_1.merge(pizzas, left_on='pizza_id',right_on='pizza_id')
4 # pizza_2.head(3)
5 pizza_4 = pizza_2.merge(pizza_types, left_on='pizza_type_id',right_on='pizza_type_id')
6 pizza_4.head(3)
7
```

จัดการ Data frame ตามความต้องการด้วย method ต่างๆ เช่น rename, drop เป็นต้น

- เปลี่ยนชื่อ column

```
1 pizza_4.rename(columns= {'date' : 'order_date',  
2                             'time' : 'order_time',  
3                             'price' : 'unit_price',  
4                             'size' : 'pizza_size'}, inplace=True)  
5
```

- ลบ column ที่ไม่ต้องการ

```
1 pizza_4 = pizza_4.drop(columns=['pizza_type_id', 'order_details_id'])
```

- สร้าง column ใหม่ตามต้องการ

```
1 pizza_4['total_price'] = pizza_4['quantity'] * pizza_4['unit_price']
```

- บันทึก data frame ที่ทำความสะอาดข้อมูลเรียบร้อยแล้วในชื่อ “pizza_transaction” เพื่อสะดวกในการทำกระบวนการถัดไป

```
1 pizza_4.to_csv('pizza_transaction.csv', index=False)
```

```
1 pizza_4.info()
```

```
>>> <class 'pandas.core.frame.DataFrame'>  
RangeIndex: 48620 entries, 0 to 48619  
Data columns (total 11 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   order_id        48620 non-null  int64  
1   pizza_id        48620 non-null  object  
2   order_date      48620 non-null  object  
3   order_time      48620 non-null  object  
4   name            48620 non-null  object  
5   pizza_size      48620 non-null  object  
6   unit_price      48620 non-null  float64  
7   quantity        48620 non-null  int64  
8   total_price     48620 non-null  float64  
9   category        48620 non-null  object  
10  ingredients      48620 non-null  object  
dtypes: float64(2), int64(2), object(7)  
memory usage: 4.1+ MB
```

วิธีการดำเนินการ : 2. ทำ Visualization เพื่อหาคำตอบหรือ Insight ตามสมมติฐานที่คาดการณ์ไว้
ซึ่งนำเนินการตามไฟล์งานที่ 2 : “ PizzaSale_Analysis ” โดยมีหลักการโดยสรุปดังนี้

สรุปภาพรวมของร้านขาย pizza

- จากข้อมูลภาพรวมร้าน pizza ปี 2015 ที่แสดงใน card chart ทำให้ได้ information ดังนี้
 - ยอดขายทั้งปีรวมเป็นเงิน 817,860 ดอลลาร์
 - ยอดขายเฉลี่ยเป็นเงิน 68,155 ดอลลาร์
 - จำนวนยอดสั่งซื้อรวมทั้งปี 21,350 ครั้ง
 - จำนวน pizza ที่ขายทั้งปี 49,574 ถาด
 - ร้านค้าขาย pizza ทั้งหมด 4 ประเภท
 - ร้านค้ามีเมนู pizza ทั้งหมด 32 เมนู

Card Chart : 2015 Pizza Store Overview

Total Revenue 817,860 \$	Total Orders 21,350	Pizza Category 4
Average Sales per Month 68,155 \$	Yearly Pizza Volume 49,574	Pizza Menu 32

ตั้งคำถามเพื่อหาคำตอบ พร้อมมองหา insight

Question 1: ยอดขายร้าน pizza น่าจะสูงที่สุดในเดือนธันวาคมใช่หรือไม่

สาเหตุที่ผู้จัดทำคาดการณ์ว่าเดือน ธันวาคม น่าจะเป็นเดือนที่ร้าน Pizza น่าจะมีรายได้สูงที่สุด
เนื่องจากมองว่าเป็นเดือนสิ้นปีที่มีจะมีการเฉลิมฉลองกับเทศกาลต้อนรับปีใหม่ น่าจะส่งผลต่อการจับจ่ายใช้
สอยของผู้บริโภค จากกราฟที่ 1 : พบว่ายอดขายที่สูงที่สุดอยู่ในเดือนกรกฎาคม ซึ่งไม่สอดคล้องกับคำถามที่
คาดการณ์เอาไว้ จากการวิเคราะห์ในเบื้องต้น(ซึ่งเป็นความเห็นของผู้จัดทำเอง) มีความเป็นไปได้หลายประการ
ที่จำทำให้เกิดเหตุการณ์เช่นนี้ อาทิเช่น

1. มีความเป็นไปได้ร้านขาย pizza เหล่งนี้จะมียอดขายช่วงเดือนธันวาคมต่ำกว่าเดือนกรกฎาคม เนื่องจากถูกแย่งลูกค้าไปเนื่องจากบริษัทคู่แข่งอาจออกโปรโมชั่นที่ดึงดูดลูกค้ามากกว่า ทำให้ลูกค้าไปซื้อสินค้าจากบริษัทคู่แข่งมากกว่า
2. ในปีนี้อาจมีสภาพอากาศที่เลวร้าย เช่น ฝนตกหนักหรืออากาศหนาวจัด อาจทำให้ลูกค้าไม่อยากจะออกจากบ้านมาซื้อ pizza

จากกราฟที่ 1 : ไม่ได้ให้ข้อมูลเราเพียงแค่เดือนธันวาคมมียอดขายที่ต่ำกว่าเดือนกรกฎาคม เท่านั้น แต่ยังทำให้เราเห็นข้อมูลรายได้ในเดือนที่ต่ำสุด และแสดงให้เห็นถึงข้อมูลยอดขายในแต่ละเดือน รวมถึงพอจะเปรียบเทียบคร่าวๆได้ว่ายอดขายเดือนไหนสูงกว่าเดือนไหนเป็นต้น

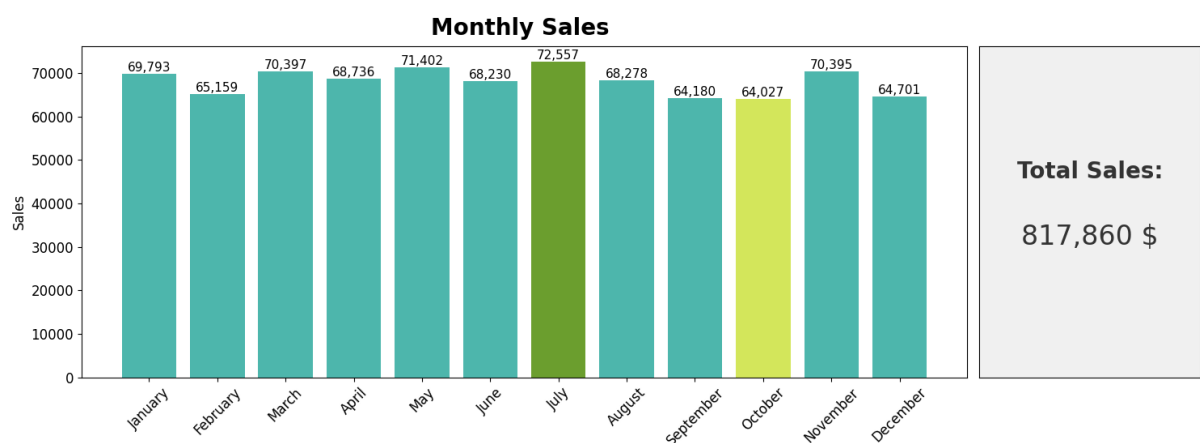


Chart no.1 : Sales by Month in 2015

จากกราฟที่ 2 : เป็นกราฟที่แสดงข้อมูลยอดขายสั่งซื้อ pizza ที่ขายได้ในแต่ละเดือนซึ่งสอดคล้องกับกราฟยอดขายตามที่แสดงไปก่อนหน้านี้

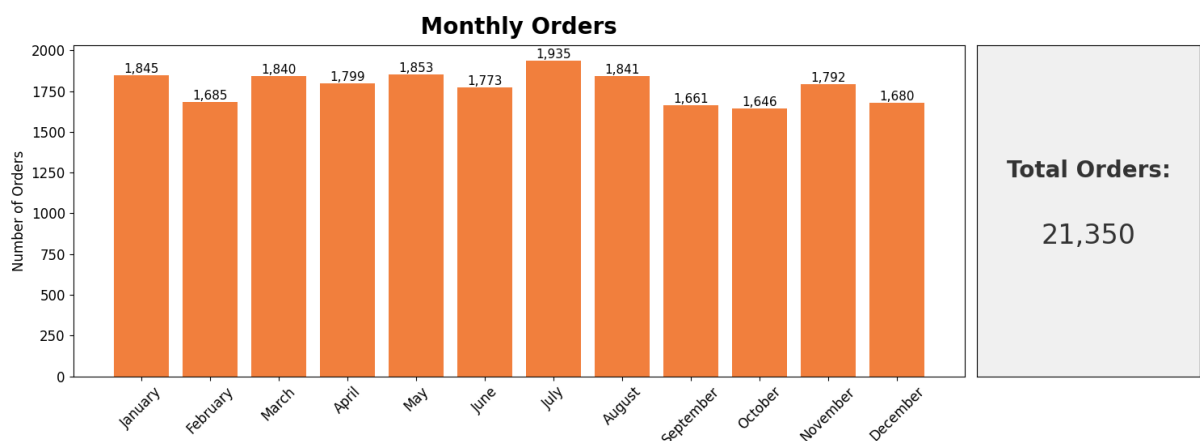
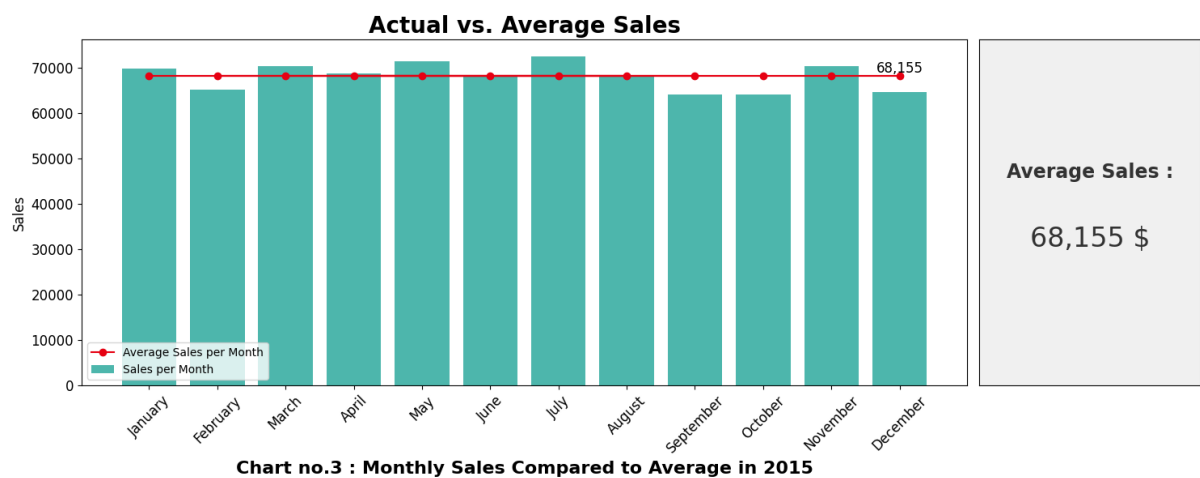


Chart no.2 : Orders by Month in 2015

จากกราฟที่ 3 : เป็นกราฟที่แสดง insight เพิ่มเติม โดยกราฟเป็นกราฟแสดงยอดขายในแต่ละเดือน เทียบกับยอดขายเฉลี่ย พบว่ายอดขายเฉลี่ยของร้าน pizza แห่งนี้อยู่ที่ 68,155 ดอลลาร์ และยังพบอีกว่าโดยส่วนใหญ่ 8 เดือนใน 12 เดือน $\left(\frac{2}{3}\right)$ ร้านค้ามียอดขายอยู่ประมาณค่าเฉลี่ย และมีเพียง 4 เดือนใน 12 เดือน $\left(\frac{1}{3}\right)$ เท่านั้นที่มียอดขายต่ำกว่าค่าเฉลี่ย ซึ่ง insight ดังกล่าว สามารถนำไปใช้ต่อยอดในพัฒนายอดขายหรือ การทำโปรโมชั่น หรืออื่นต่อไปได้ ยกตัวอย่างเช่น เราสามารถนำ insight ที่ได้นี้ไปใช้ประโยชน์ในการเลือก เดือนที่จะทำโปรโมชั่นในการกระตุ้นยอดขาย โดยอาจจะเลือกจากเดือนที่มียอดขายต่ำ หรือนำข้อมูลดังกล่าว ไปวิเคราะห์หาสาเหตุที่แท้จริงที่ทำให้เดือนดังกล่าวมียอดขายที่ต่ำกว่าค่าเฉลี่ย เป็นต้น



Question 2: ยอดขายร้านน่าจะมาจาก pizza ประเภท Classic มากที่สุดใช่หรือไม่

จากกราฟที่ 4 : จะเห็นว่ายอดขาย pizza ในประเภท Classic นั้นสูงที่สุดคิดเป็น 27% จากยอดขายทั้งหมด รองลงมาเป็นประเภท Supreme คิดเป็น 25% จากยอดขายทั้งหมด สำหรับประเภท Chicken และ Veggie มีรายได้คิดเป็น 24% จากยอดขายทั้งหมด โดยภาพรวมแล้วยอดขายของแต่ละประเภทดูไม่มีความแตกต่างกันมากนัก

ทั้งนี้จากกราฟที่ 5 : สามารถบอกข้อมูลในเชิงลึกได้ว่ายอดขาย 27% จากยอดขายทั้งหมด คิดเป็นจำนวนเงินเท่ากับ 220,053\$ สำหรับยอดขาย pizza ในประเภท Classic และสามารถดูข้อมูลในเชิงลึกนี้ได้กับทุกประเภทของ pizza

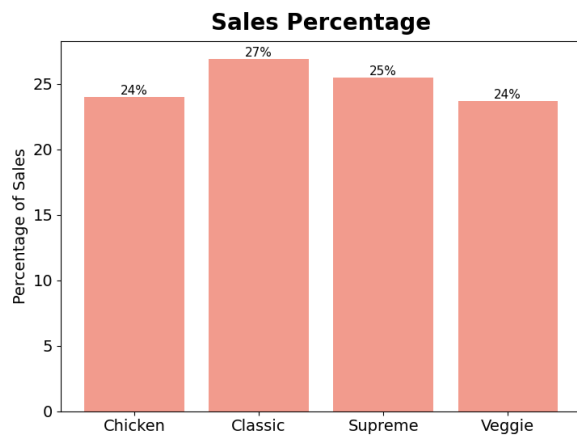


Chart no.4 : Sales Percentage by Category

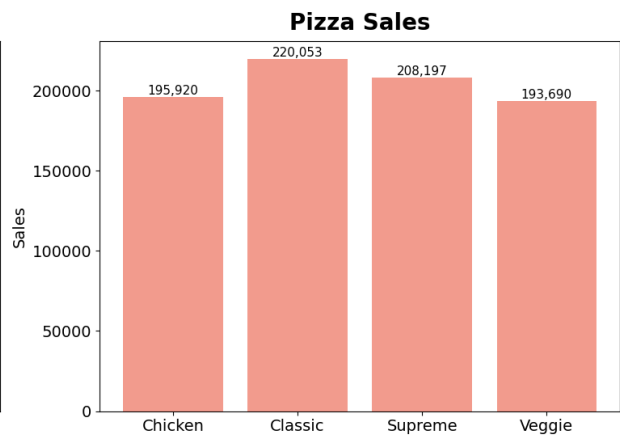


Chart no.5 : Sales Analysis by Category

จากกราฟที่ 6 : แสดงจำนวน pizza ที่ขายได้แยกตามประเภท พบว่า Pizza ประเภท Classic มีจำนวน pizza ที่ขายได้มากที่สุดโดยคิดเป็นจำนวน 14,579 ถาด, รองลงมาเป็น pizza ประเภท Supreme คิดเป็นจำนวน 11,777 ถาด, pizza ประเภท Chicken คิดเป็นจำนวน 11,449 ถาด และ pizza ประเภท Veggie เป็นจำนวน 10,815 ถาด ตามลำดับ

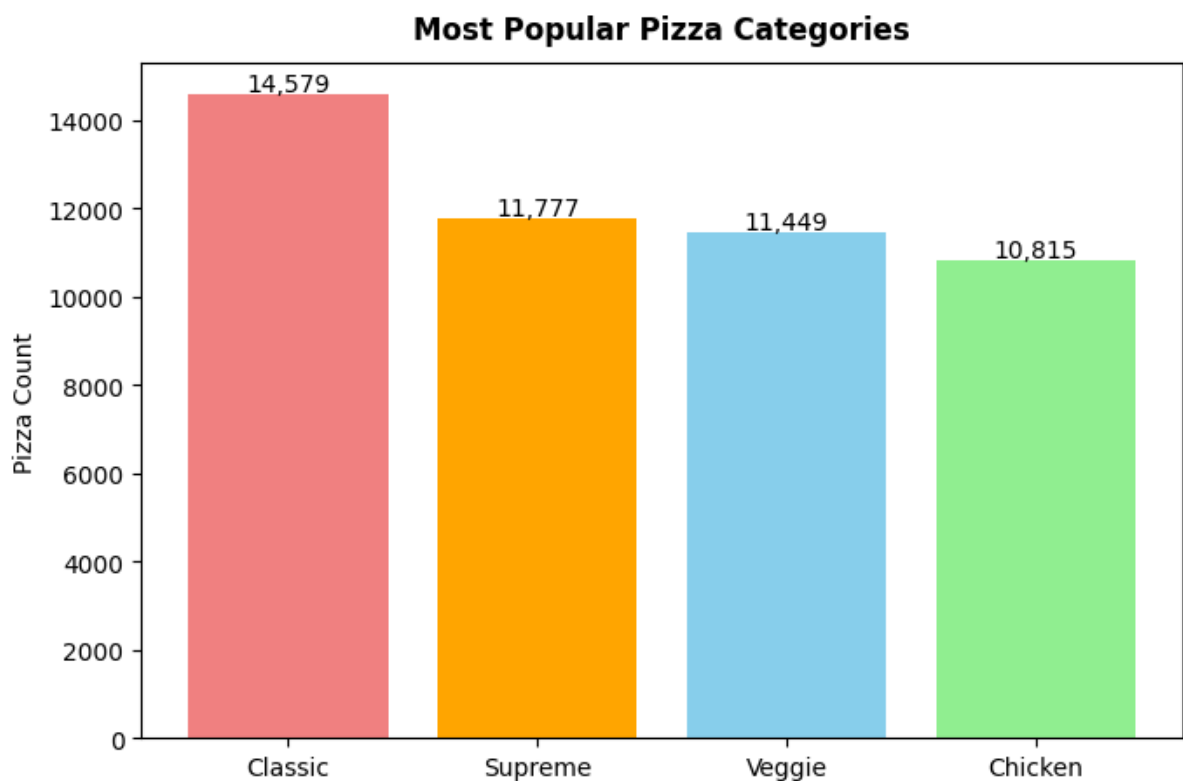


Chart no.6 : Top Pizza Categories

Question 3: Hawaiian pizza น่าจะเป็นหนึ่งในเมนูยอดนิยมใช่หรือไม่

จากกราฟที่ 7 : แสดงจำนวน pizza ที่ขายได้สูงสุด 5 อันดับแรก ดังนี้

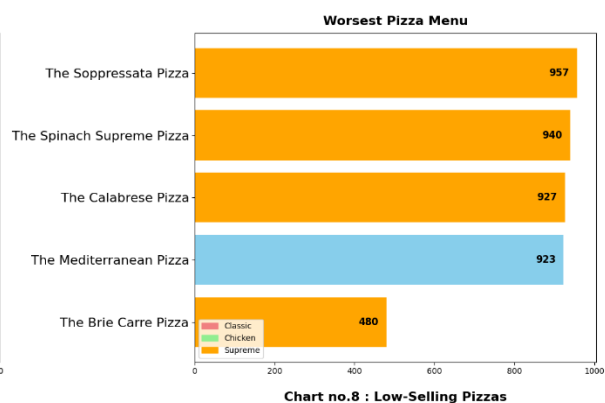
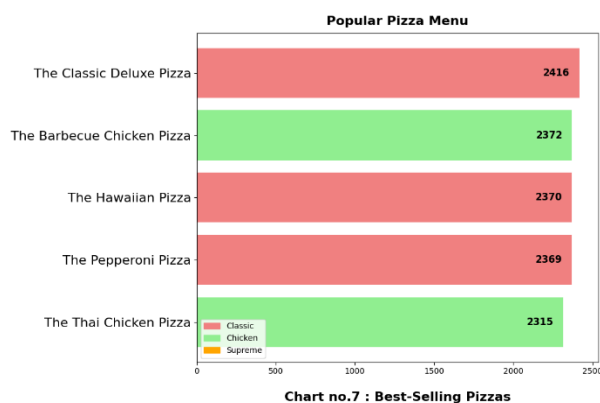
1. The Classic Deluxe Pizza ตลอดทั้งปีขายได้เป็นจำนวน 2,416 ถาด
2. The Barbecue Chicken Pizza ตลอดทั้งปีขายได้เป็นจำนวน 2,372 ถาด
3. The Hawaiian Pizza ตลอดทั้งปีขายได้เป็นจำนวน 2,370 ถาด
4. The Pepperoni Pizza ตลอดทั้งปีขายได้เป็นจำนวน 2,369 ถาด
5. The Thai Chicken Pizza ตลอดทั้งปีขายได้เป็นจำนวน 2,315 ถาด

ซึ่งแสดงให้เห็นว่า Hawaiian Pizza เป็น 1 ใน 5 อันดับของ pizza ที่ขายได้สูงสุด จึงเป็นหนึ่งในเมนูยอดนิยมของลูกค้าในปีนี้ และสามารถใช้ข้อมูลในการจัดทำเมนูแนะนำของร้านค้าได้

จากกราฟที่ 8 : แสดงจำนวน pizza ที่ขายได้ต่ำสุด 5 อันดับ ดังนี้

1. The Soppressata Pizza ตลอดทั้งปีขายได้เป็นจำนวน 957 ถาด
2. The Spinach Supreme Pizza ตลอดทั้งปีขายได้เป็นจำนวน 940 ถาด
3. The Calabrese Pizza ตลอดทั้งปีขายได้เป็นจำนวน 927 ถาด
4. The Mediterranean Pizza ตลอดทั้งปีขายได้เป็นจำนวน 923 ถาด
5. The Brie Carre Pizza ตลอดทั้งปีขายได้เป็นจำนวน 480 ถาด

จากข้อมูลดังกล่าวสามารถนำไปใช้ในการวิเคราะห์ในด้านคุณภาพ ราคา ความต้องการของลูกค้าหรือความล้าสมัยของ Pizza ได้



Question 4: ลูกค้าน่าจะชอบสั่งซื้อ pizza ในวันศุกร์ช่วงเวลาเย็นใช่ไหม

จากกราฟที่ 9 : พบว่าวันศุกร์เป็นวันที่ขาย pizza ได้มากที่สุด โดยตลอดทั้งปีวันศุกร์ขาย pizza ได้เป็นจำนวน 8,106 ถาด รองลงมาเป็นวันเสาร์ ขายได้จำนวน 7,355 ถาด และวันพฤหัสบดีขายได้จำนวน 7,323 ถาด ตามลำดับ

จากกราฟที่ 10 : แสดงจำนวน pizza ที่ขายได้ในวันศุกร์ตามช่วงเวลาตั้งแต่ 9 โมงเช้าไปจนถึงเที่ยงคืน พบว่ามี 2 ช่วงเวลาที่มีขาย pizza ได้เป็นจำนวนมากคือคือช่วงเที่ยงถึงบ่ายโมง [12:00PM – 13:00PM] และช่วงห้าโมงเย็นถึงหนึ่งทุ่ม [17:00PM – 19:00PM]

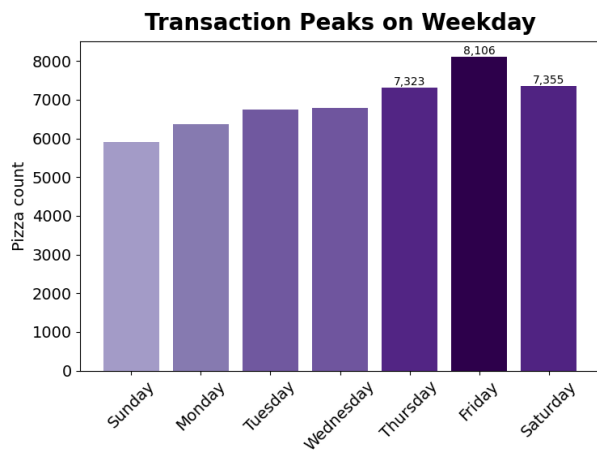


Chart no.9 : High Transaction Days

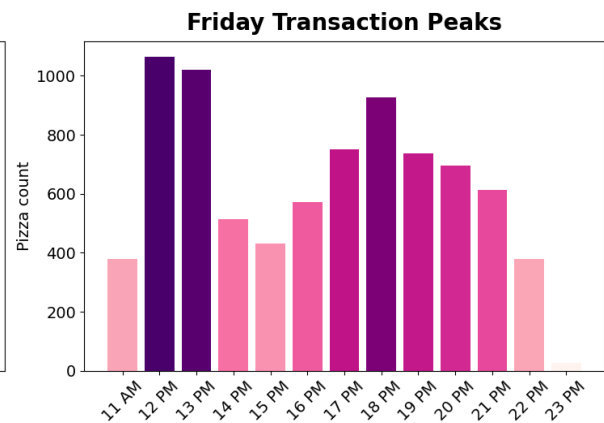


Chart no.10 : Hourly Transaction Peaks on Friday

Question 5: โดยภาพรวมลูกค้าน่าจะชอบสั่ง pizza ในช่วงเวลาเย็นใช่ไหม

จากกราฟที่ 11 : แสดงจำนวน pizza ที่ขายได้ โดยแบ่งตามช่วงเวลาตั้งแต่ 9 โมงเช้าไปจนถึงเที่ยงคืน พบว่ามี 2 ช่วงเวลาที่มียอดขายค่อนข้างสูงคือช่วงเที่ยงถึงบ่ายโมง [12:00PM – 13:00PM] และช่วงห้าโมงเย็นถึงหนึ่งทุ่ม [17:00PM – 19:00PM]

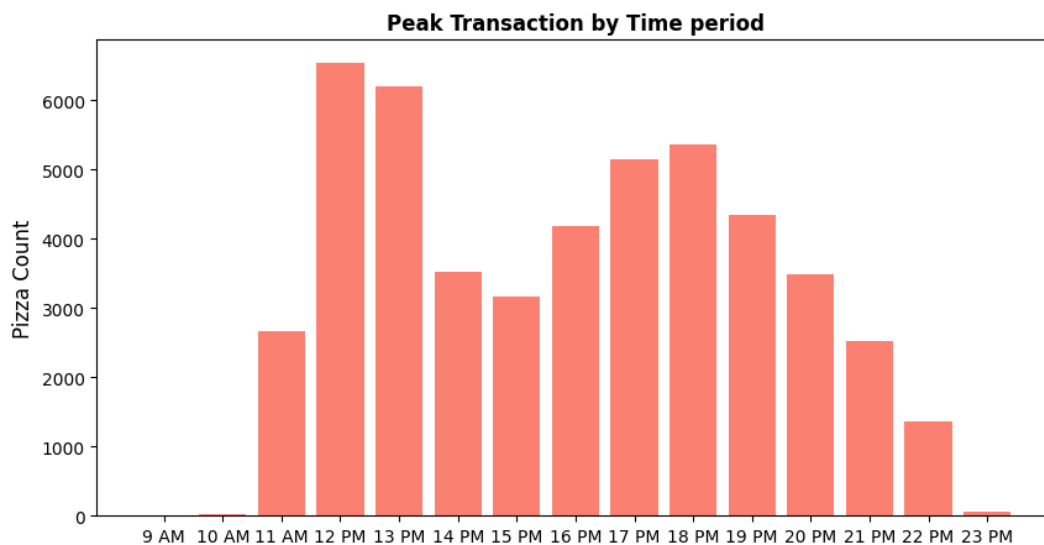


Chart no.11 : Hourly Transaction Peaks in 2015