Introduction: The Zomato data presents a wealth of information about various restaurants, their customers, and sales data. This analysis intends to focus on customer segmentation in order to determine trends in their purchasing habits based on various criteria. The main questions to be answered in this analysis:
What overall customer segments are ordering most often?
Are there certain demographics that consistently spend more than others?
How does the size of the family relate to sales?
Are there any social factors at work in terms of education or occupation?..

Data Description: The overall data consists of multiple tables that cover the users, food me..

Introduction: The Zomato dat..

Data Description: The overall data consists of multiple tables that cover the users, food menus, restaurants, and orders. For this analysis the focus is going to be on the users and orders tables so that we can understand how demographic information about our users can inform us about their ordering patterns. The Orders table contained information on the user id associated with the order, order date, the quantity of items ordered, the total order cost, currency, and restaurant id. The Users table provided detailed information on each user from user id, name, email, password, age, occupation, education level, family size, income group, gender, and marital status. The demographic information provided was a mixture of text and numerical data, while the orders data consisted of date time and numerical data.

Methodology: Data preparatio..

	user_id	name	email	password	Age	Gender	Marital Statu	Occupation	Monthly Incc	Educational	Family size
0	1	Claire Ferguson	fordanthony@ex	NKz0fWDh!5	20	Female	Single	Student	No Income	Post Graduate	4
1	2	Jennifer Young	ann96@example	+i5Q91jt!s	24	Female	Single	Student	Below Rs.10000	Graduate	3
2	3	Jermaine Robers	uwalker@examp	eO4GqGusF(	22	Male	Single	Student	Below Rs.10000	Post Graduate	3
3	4	Rachel Carpente	kimberlypatterso	d^c7O6odaS	22	Female	Single	Student	No Income	Graduate	6
4	5	Shawn Parker	daniellebennett@	8J#E5RMl1o	22	Male	Single	Student	Below Rs.10000	Post Graduate	4
5	6	Timothy Clark	brettsantana@ex	qOJ0NA0Y$4	27	Female	Married	Employee	More than 50000	Post Graduate	2
6	7	Alexander Lucas	susan58@exam	(^+21Yv3Uv	22	Male	Single	Student	No Income	Graduate	3
7	8	Christopher Curr	brookesmith@ex	^+5fP5zm(L	24	Female	Single	Student	No Income	Post Graduate	3
8	9	Daniel Mercado	imyers@example	e$DJ2tRoJ#	23	Female	Single	Student	No Income	Post Graduate	2
9	10	Tony Lawrence	lwelch@example	&3emljZp#O	23	Female	Single	Student	No Income	Post Graduate	4
10	11	Scott Sherman	brianbaird@exar	0hJuTqHk!8	22	Female	Single	Student	No Income	Post Graduate	5
11	12	Michael Gilbert	wendycollins@ex	oF1Yyh4c@I	23	Male	Single	Student	Below Rs.10000	Post Graduate	2
12	13	Alex Cortez	jennifer67@exan	*4XZKtvji							
13	14	Micheal King	william41@exam	_QcC+7Z							
14	15	Daniel Lopez	marygarcia@exa	0gXloziG							
15	16	Carla Frye	tamara97@exam	&k8F)Kwt							
16	17	Theresa Murphy	chelseajordan@	%GHX31							
17	18	Antonio Brown	meyernicole@ex	#Q1mKck							
18	19	Kyle Hall	laurie00@examp	lc8B7!uR							
19	20	Adrienne Davids	phillipskevin@ex	^7NF^tug							
20	21	Chris Gonzalez	wmeyers@exam	yEa5GD1							
21	22	Patrick Ibarra	melissahopkins@	k%4V1TH							

	order_date	sales_qty	sales_amount	currency	user_id	r_id
0	2017-10-10	100	41241	INR	49226	567335
1	2018-05-08	3	-1	INR	77359	531342
2	2018-04-06	1	875	INR	5321	158203
3	2018-04-11	1	583	INR	21343	187912
4	2018-06-18	6	7176	INR	75378	543530
5	2017-11-20	59	500	USD	34323	158204
6	2017-11-22	36	250	USD	33246	156588
7	2017-11-23	39	21412	INR	87420	244866
8	2017-11-27	35	19213	INR	31017	156602
9	2017-11-28	310	170185	INR	72391	158193
10	2017-11-29	184	101194	INR	91457	407249
11	2017-11-30	35	19213	INR	33851	156590
12	2017-11-29	17	9426	INR	27008	338749
13	2017-12-19	1	218	INR	47798	156601
14	2018-08-07	5	3093	INR	58443	156587
15	2017-12-04	58	30306	INR	32969	158192

Data Description: The overal..	Methodology: Data preparation began with cleaning and validating the data in the Orders and Users tables. Sales amounts were converted to US Dollars from Indian Rupees based on current conversion rates (1 INR = 0.012065 USD). The same conversion was used on the represented income groups to convert to USD ranges. This will allow us to better interpret the results of our analyses and understand the implications. Furthermore, when looking at the sales data any blank, zero, or negative values were removed so that only actual sales/orders were being represented. Data cleaning then led to the removal of superfluous column information like User passwords and emails and the restaurant ids from the Orders table. The two tables were then joined based on the user id column so that analysis could begin. Some initial exploratory data analysis in Excel and SQL looked at demographic rates to show that certain demographics were much more prevalent and that there were some extremely high sales values coming from unexpected sources. As a result, the decision ..	Analysis: Initially, it was ..

```sql
SELECT DISTINCT users.name, orders.currency_usd as Sales_USD, orders.sales_qty as Quantity, users.Incomegroup, users.Familysize, users.Occupation, COUNT(orders.order_date) as Num_Orders
FROM orders
JOIN users ON users.user_id = orders.user_id
GROUP BY users.user_id
ORDER BY orders.sales_amount DESC, Num_Orders DESC
LIMIT 25;
```

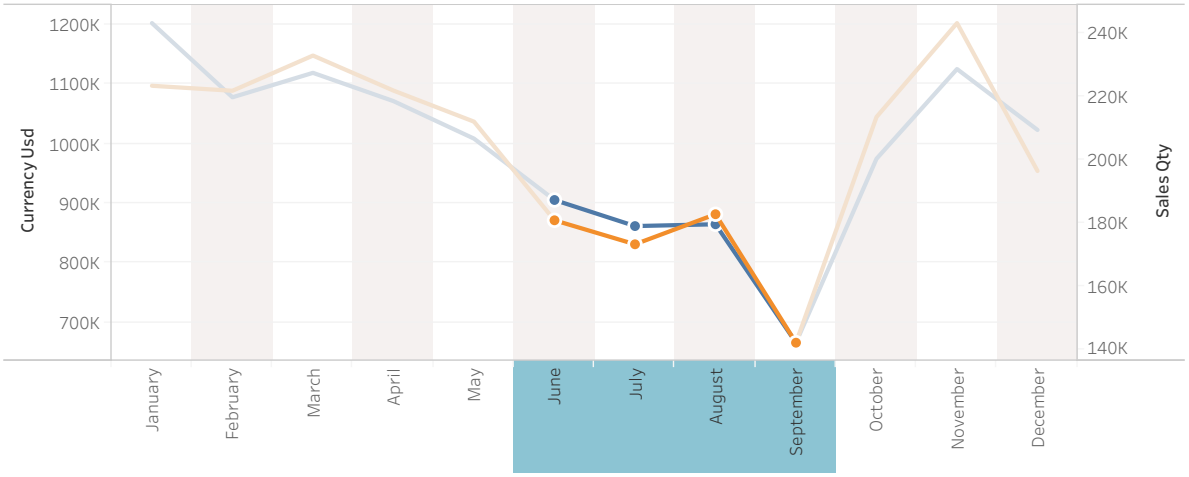
	name	Sales_USD	Quantity	Incomegroup	Familysize	Occupation	Num_Orders
1	Amanda Ballard	$18,229.54	725	4	2	Self Employeed	3
2	Richard Edwards	$16,146.16	1798	3	3	Employee	1
3	Cassandra Benson	$15,489.95	600	3	2	Employee	3
4	Brian White	$15,489.95	600	1	2	Student	3
5	Elizabeth Ryan	$14,904.46	587	3	4	Employee	2
6	Jeffrey Smith	$14,817.61	1480	2	5	Student	1
7	Billy Wilson	$14,768.62	1120	0	2	Student	3
8	Melissa Brown	$14,004.83	280	1	3	Student	2
9	Kevin Thomas	$12,161.18	490	4	3	Self Employeed	1
10	Jeremy Richardson	$12,017.97	747	3	5	Employee	4
11	Kristina Powers	$11,995.52	376	0	3	Student	3
12	Elizabeth Wells	$11,704.62	778	0	4	Student	2
13	Brittany Mcfarland	$11,704.62	778	4	2	Employee	2
14	Joshua Rice	$10,379.59	490	4	4	Employee	1
15	Maria Harris	$10,049.19	680	4	2	Self Employeed	1
16	Lisa Jones	$9,740.09	400	3	3	Employee	2
17	Diana Chandler	$9,393.66	393	3	3	Self Employeed	2
18	Derrick Carter	$8,755.56	587	4	4	Employee	3
19	Eric Henry IV	$8,747.57	443	2	3	Self Employeed	1
20	Jordan Parker	$8,046.18	343	0	2	Student	1
21	Maria Daniel	$8,036.63	800	0	2	Student	3
22	James Diaz	$7,955.42	320	4	6	Self Employeed	2
23	Dana Gonzalez	$7,955.42	320	0	5	Student	2
24	Barbara Johnston	$7,922.46	160	2	5	Self Employeed	1

Meth odolo gy: D ata pr epara tio..

Analysis: Initially, it was important to get an understanding of the general sales trends on the app to see what kind of trends or patterns stood out. Naturally, the amount of items being sold was closely aligned to our overall sales amounts through the year. Right away there's a noticeable decline in the amount of sales as we head into June throughout September. This is an alarming dropoff in profitability that needs to be understood in order to be better addressed. When looking at how different demographic segments show across their total sales, number of orders, and median sales amounts certain categories tend to stand out as the biggest app users. Across the different demographic segments: Males, Single individuals, Graduate and Post Graduate level educated, Students and Employees, and households made up of 2 to 3 people make up the bulk of users and spenders. Regardless of how big of users the different groups are we still see similar declines in the summer months in the number of orders and overall sales; however, median sales amounts tend ..

In this view we can better analyze the breakd..

Sales Trends by Month



Measure Names
Currency Usd
Sales Qty

Demographics 1
Occupation

Calculations
Number_of_Orders

Calculations
218 7,633

Monthly Number_of_Orders by Occupation

Analysis: Initially, it was ..

In this view we can better analyze the breakdown of different demographic categories and what percent of the total user population they make up. In the heatmap below we can use the same view and characteristics but see where different calculations show up as a result. The general outcome is the same correlation where the more overall representation a segment has in the population the greater the raw amount of sales and number of orders. What's unique is when we compare the average number of items being bought and median sales amounts. For instance, let's look below at a comparison between genders and their occupations. While the majority of males and females are students, we see that Female employees have higher median sales than their student counterparts, while male students still show higher median sales.

When duplicating the same de..

Demographic Breakdowns

Demographics 2	Employee	House wife	Self Employeed	Student
Female	12.66%	2.31%	4.15%	23.71%
Male	17.82%		9.81%	29.55%

Demographics 1
Occupation

Demographics 2
Gender

Calculations
5.9200 6.3950

Calculations
Median Sales

Multiple Demographic Heatmaps

Demographics 2	Employee	House wife	Self Employeed	Student	Grand Total
Female	6.3950	6.2600	5.9200	6.2000	**6.2600**
Male	6.2600		6.2600	6.3700	**6.3100**
Grand Total	**6.3100**	**6.2600**	**6.1400**	**6.3100**	**6.2600**

In this view we can better a..

When duplicating the same demographic we're able to see the standouts in each category. Males buy and spend more on average than females. School educated individuals buy more but spend less than other education levels. Those that don't identify their marital status buy and spend more but also make up a much smaller percentage of the population which could signal some outliers in the data. Employed individuals (Employee or Self Employed) buy more but spend roughly the same as students. Income earners between 3600 and 7500 USD tend to spend the most. Lastly, single individuals tend to buy the most food items compared to larger families that may naturally spend more.

Now that we have a better underst..

Demographic Breakdowns

Demographics 2	1	2	3	4	5	6
1	6.23%					
2		26.10%				
3			30.05%			
4				16.19%		
5					13.90%	
6						7.53%

Multiple Demographic Heatmaps

Demographics 2	1	2	3	4	5	6	Grand T
1	20.009						
2		16.812					
3			15.545				
4				15.557			
5					16.286		
6						15.628	
Grand Total	20.009	16.812	15.545	15.557	16.286	15.628	

Demographics 1
Family size

Demographics 2
Family size

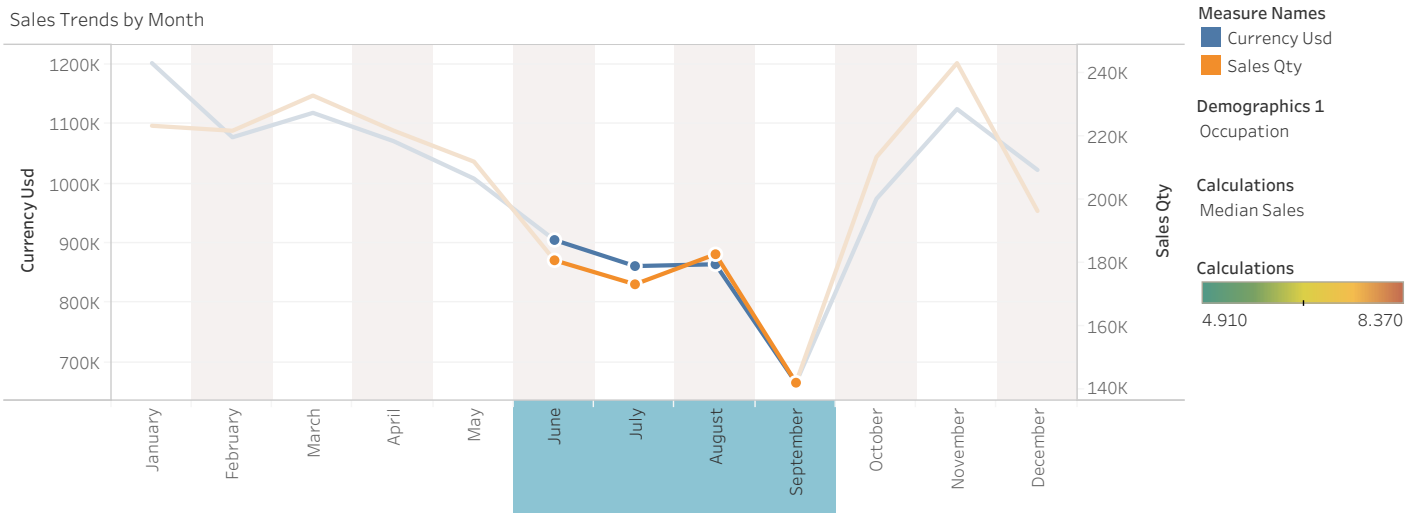
Calculations
15.545 20.009

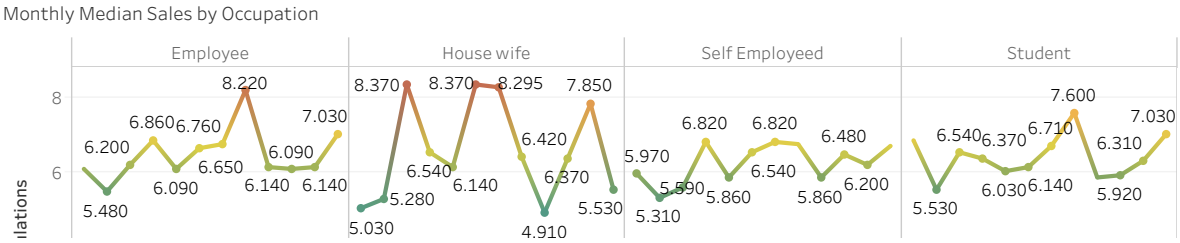
Calculations
Avg Quantity

When duplicating the same de..

Now that we have a better understanding of which demographic segments tend to be most profitable and bigger users of the app this raises an interesting question that was not initially considered when planning the analysis, which is, what is the most profitable time frame and how can we address the less profitable summer months?

When researching potential c..



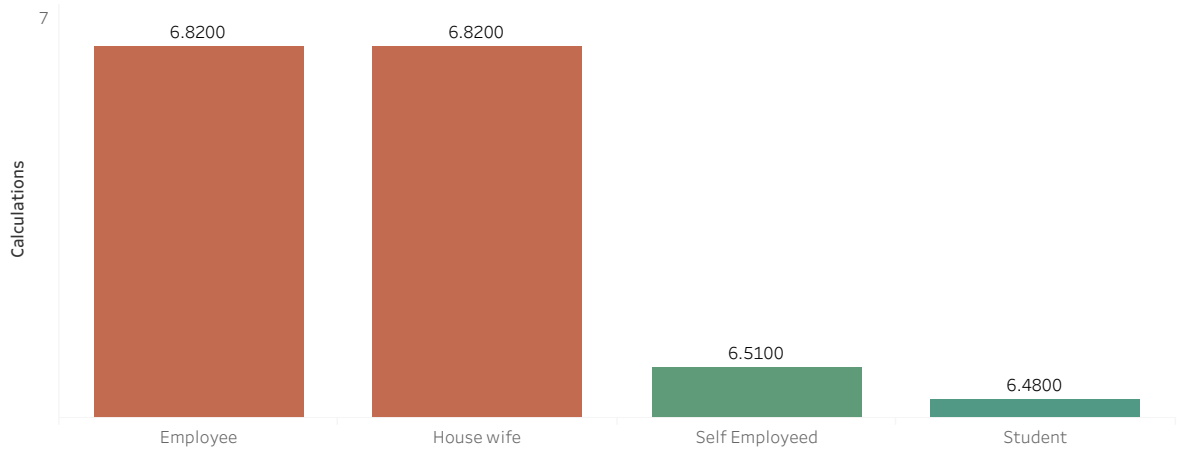


Now that we have better u n..

When researching potential causes for large drops in the revenue of the app, a couple of options came up. For one, given that a large proportion of our users are students, perhaps the school year was resposible for the drop in sales. However, that wouldn't account for the drop in sales that we also see across the other segments. What actually seems most likely is the corresponding monsoon season in India that runs from June through September. Sales drop precipitously during this frame but median sales amounts actually tend to jump up as seen below. Perhaps, users are stocking up on food so as to not have to go out as frequently, while there may also be less availability of certain items from the restaurants.

Conclusion: Through our anal..

Monsoon Season Sales



Calculations
6 7

Demographics 1
Occupation

Calculations
Median Sales

Non Monsoon Season Sales

When researching potential causes for large drops in the revenue of the app, a couple of options came..

Conclusion: Through our analysis we can see that the largest populations of our user group fall into the following categories: single, male, students, with graduate and post graduate education levels. Our marketing strategies should not only continue to encourage these groups to use the app for their needs but also look to expand into these other under utilized markets. Generally, women, housewives and employees, married individuals, and families with more than 2 people are spending in line or even greater on average with the larger user populations and represent an untapped market. We should focus on showing our value of giving back more time to the family, less stress, and more time to work on their own hobbies and passions as ways to bring them into the fold. We could also look to incentivize these groups to start using the app and to keep them coming back. This would also be true during the monsoon season. Focusing on ways to get overall order numbers up even close to the non monsoon season will help to make our profitability more sustainable throu..