

# Comparing Customer Satisfaction Across Different Regions

Use a one-way ANOVA test to compare customer satisfaction scores across different geographical regions identified in the dataset.

**$H_0$**

**There is no significant difference in customer satisfaction scores  
across different geographical regions**

---

**$H_1$**

**There is no significant difference in customer satisfaction scores  
across different geographical regions**

# Customer Personality Analysis

category	Sub-category	Customer Remarks	Order_id	order_date_time	Issue_reported at	issue_responded	Survey_response_Date	Customer_City	connected_handling_time	Agent_name	Supervisor	Manager	Tenure Bucket	Agent Shift	CSAT Score
Product Queries	Life Insurance	NaN	c27c9bb4-fa36-4140-9f1f-21009254ffdb	NaN	01/08/2023 11:13	01/08/2023 11:47	01-Aug-23	NaN	NaN	Richard Buchanan	Mason Gupta	Jennifer Nguyen	On Job Training	Morning	5
Product Queries	Product Specific Information	NaN	d406b0c7-ce17-4654-b9de-f08d421254bd	NaN	01/08/2023 12:52	01/08/2023 12:54	01-Aug-23	NaN	NaN	Vicki Collins	Dylan Kim	Michael Lee	>90	Morning	5
Order Related	Installation/demo	NaN	c273368d-b961-44cb-beaf-62d6fd6c00d5	NaN	01/08/2023 20:16	01/08/2023 20:38	01-Aug-23	NaN	NaN	Duane Norman	Jackson Park	William Kim	On Job Training	Evening	5
Returns	Reverse Pickup Enquiry	NaN	5aed0059-55a4-4ec6-bb54-97942092020a	NaN	01/08/2023 20:56	01/08/2023 21:16	01-Aug-23	NaN	NaN	Patrick Flores	Olivia Wang	John Smith	>90	Evening	5

# Variable Selection

---

**Independent Variable:** Customer\_City

**Dependent Variable:** cat\_score

# Customer Satisfaction Analysis

```
cities = data["Customer_City"].value_counts()
cities

Customer_City
HYDERABAD    722
NEW DELHI    688
PUNE         435
MUMBAI       406
BANGALORE    352
...
GUNTAKAL     1
MANSAR       1
BAGHMARA     1
HINDORIA     1
DORAHA       1
Name: count, Length: 1782, dtype: int64

top_cities = cities[cities.values > 400].index.tolist()
top_cities

['HYDERABAD', 'NEW DELHI', 'PUNE', 'MUMBAI']
```

# Customer Satisfaction Analysis

---

```
data["CSAT Score"].describe()
```

```
count      85907.000000  
mean        4.242157  
std         1.378903  
min         1.000000  
25%         4.000000  
50%         5.000000  
75%         5.000000  
max         5.000000  
Name: CSAT Score, dtype: float64
```

# Customer Satisfaction Analysis

```
categories = data.groupby("category").groups.keys()
print("Total no of Groups: ", data.groupby("category").ngroups)
categories
```

Total no of Groups: 12

```
dict_keys(['App/website', 'Cancellation', 'Feedback', 'Offers & Cashback', 'Onboarding related', 'Order Related', 'Others', 'Payments related', 'Product Queries', 'Refund Related', 'Returns', 'Shopzilla Related'])
```

```
dict1 = {}
for category in categories:
    dict1[category] = random.uniform(0, 1)
```

dict1

```
{'App/website': 0.5937642826000437,
 'Cancellation': 0.2574987290096903,
 'Feedback': 0.8286538956845502,
 'Offers & Cashback': 0.02046915521766013,
 'Onboarding related': 0.9632905624293855,
 'Order Related': 0.5756678521065155,
 'Others': 0.7195585522159607,
 'Payments related': 0.7472204310770725,
 'Product Queries': 0.8091689172057871,
 'Refund Related': 0.05581143801753485,
 'Returns': 0.6382394963391267,
 'Shopzilla Related': 0.3468240414977014}
```

# Customer Satisfaction Analysis

```
merged_data["SAT SCORE"] = merged_data["CSAT Score"] * merged_data["SAT_Score"]  
merged_data.columns
```

	category	SAT_Score
0	App/website	0.593764
1	Cancellation	0.257499
2	Feedback	0.828654
3	Offers & Cashback	0.020469
4	Onboarding related	0.963291
5	Order Related	0.575668
6	Others	0.719559
7	Payments related	0.747220
8	Product Queries	0.809169
9	Refund Related	0.055811
10	Returns	0.638239
11	Shopzilla Related	0.346824



	Customer_City	SAT_Score
0	NEW DELHI	0.575668
1	NEW DELHI	0.638239
2	PUNE	0.638239
3	PUNE	0.257499
4	NEW DELHI	0.638239
...	...	...
2246	MUMBAI	0.575668
2247	HYDERABAD	0.575668
2248	MUMBAI	0.575668
2249	PUNE	0.638239
2250	PUNE	0.575668

```
merged_data.describe()
```

	SAT_Score
count	2251.000000
mean	0.556527
std	0.161284
min	0.020469
25%	0.575668
50%	0.575668
75%	0.638239
max	0.828654



# Customer Satisfaction Analysis

```
one_way = final.groupby("Customer_City")["SAT_Score"]  
one_way.ngroups
```

4

```
for city, group in one_way:  
    print(city + " : \n" + str(group))  
    print()
```

HYDERABAD :

5	0.747220
6	0.575668
13	0.257499
14	0.638239
23	0.828654
90	0.055811
1730	0.020469
1790	0.719559

Name: SAT\_Score, dtype: float64

MUMBAI :

8	0.575668
10	0.638239
24	0.257499
38	0.828654
102	0.346824
166	0.055811
218	0.020469

Name: SAT\_Score, dtype: float64

# Customer Satisfaction Analysis

---

```
city_groups = []  
for city, group in one_way:  
    city_groups.append(group.tolist())
```

```
f_statistic, p_value = f_oneway(*city_groups)
```

```
print("F-statistic:", f_statistic)  
print("p-value:", p_value)
```

```
F-statistic: 0.20481836264598535  
p-value: 0.8921201040656405
```

# Customer Satisfaction Analysis

---

```
if p_value < 0.05:
    print("""
    1. Data is statistically significant
    2. We have evidence to support the alternative hypothesis
    3. Null Hypothesis is Rejected.

    Conclusion:
        There is a significant difference in customer satisfaction scores across different geographical regions.
    """)
else:
    print("""
    1. Data is statistically not significant
    2. Null Hypothesis is Accepted.

    Conclusion:
        There is no significant difference in customer satisfaction scores across different geographical regions.
    """)
```

1. Data is statistically not significant
2. Null Hypothesis is Accepted.

Conclusion:  
There is no significant difference in customer satisfaction scores across different geographical regions.

# Conclusion

$H_0$

There is no significant difference in customer satisfaction scores across different geographical regions

---

Null Hypothesis Accepted

**Thank You !**