

E-Commerce: Seasonal and Geographic Trends





INTRODUCTION

**What is the difference between
traditional retailing and E-commerce?**

Traditional Retail

Customers located nearby

Customers in the same season

E-commerce

Customers located worldwide

Customers at different seasons



Geolocation & Seasonality





Orders Cancellation

Inventory Failure

- Fail to account for seasonal product
- Fail to account for high demand area

Delivery Failure

- Fail to account for bad weather
- Fail to account for certain area's transportation condition





LITERATURE REVIEW (Seasonality)

Demand seasonality in retail inventory management

Ehrenthal, J. C. F., Honhon, D., & Van Woensel, T. (2014).

- Not accounting for demand seasonality can lead to substantial optimality gaps, yet incorporating only some form of demand seasonality does not always lead to cost savings

Seasonal relevance in e-commerce search

Yang, H., Gupta, P., Fernández Galán, R., Bu, D., & Jia, D. (2021).

- The treatment with seasonal relevance features results in 2.20% higher purchases and better customer experience overall



LITERATURE REVIEW (Geolocation)



E-Commerce, Transportation, and Economic Geography

Anderson, W. P., Chatterjee, L., & Lakshmanan, T. R. (2003).

- A fast penetration of online shopping can be expected in more remote areas where e-commerce improves retail accessibility / Early e-shopping to be limited to urban areas

Identifying the geography of online shopping adoption in Belgium

Beckers, J., Cárdenas, I., & Verhetsel, A. (2018).

- The demand for e-commerce deliveries is not evenly distributed over the population



PROJECT HYPOTHESIS



H₀

The combination of season and geolocation
has **no significant effect** on cancellation

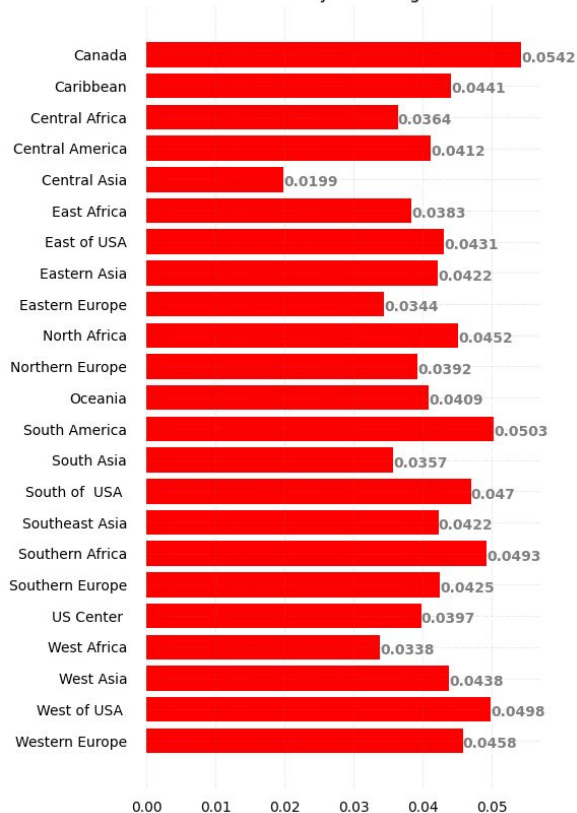
H_a

The combination of season and geolocation
has **significant effect** on cancellation

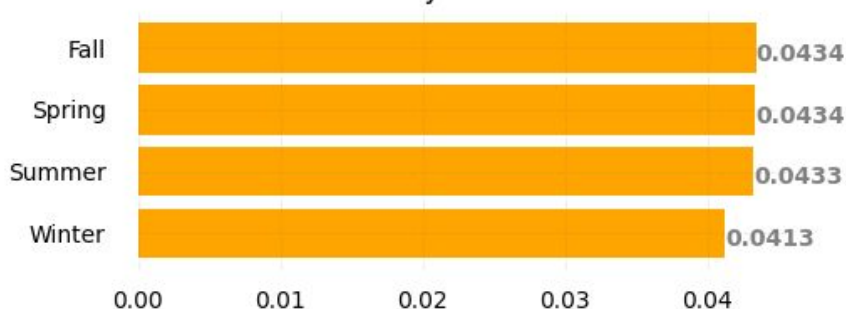
DESCRIPTIVE STATISTICS



Cancellation Rates by Order Region



Cancellation Rates by Season



“Seasons” added to the dataset

- Values were populated in the dataset by building a function to determine the season for a given month and region based on which hemisphere the area falls in

Calculated average cancellation rates per order region and season

- Preliminary results show that seasons have less effect on the cancellation rates.



Statistical Models



01

ANOVA

- how 'Season' and 'Order Region' affect 'Cancellation Rate'



02

Multiple Linear Regression

- H0: the combination of seasonality and geographic location has **no significant influence** on cancellation rates
- Ha: the combination of seasonality and geographic location has **significant influence** on cancellation rates

03

Logistic Regression

- H0: the combination of seasonality and geographic location has **no significant influence** on order cancellations
- Ha: the combination of seasonality and geographic location has **significant influence** on order cancellations

04

Clustering

- Categorical variables are divided into clusters based on the cancellation rate for division.

01. ANOVA



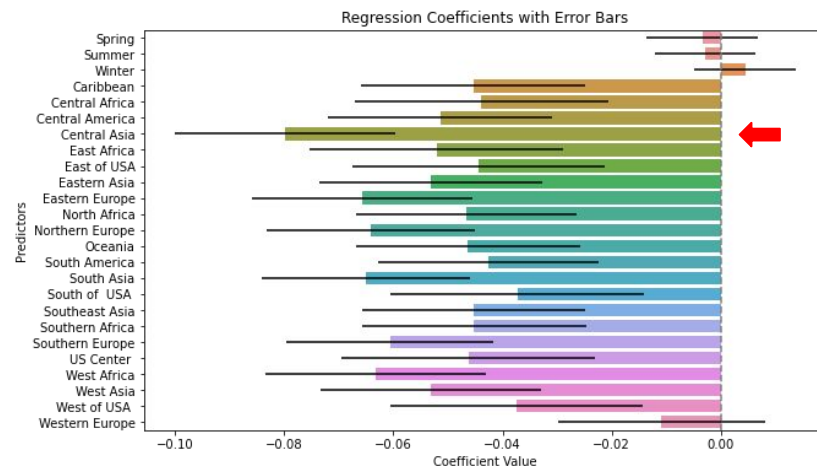
Used two-tail ANOVA to check
how 'Season' and 'Order Region' affect
'Cancellation Rate'

- Capturing any type of change, whether increase or decrease, including combined effects of these factors
- Alpha = 0.15
- Seasons variable appeared to be not significant
- Order Region appeared to be significant

	sum_sq	df	F	PR(>F)
C(Season)	0.000544	3.0	0.295909	0.828121
C(Q("Order Region"))	0.020710	22.0	1.535969	0.115590
Residual	0.025128	41.0	NaN	NaN

02. Multi-linear Regression

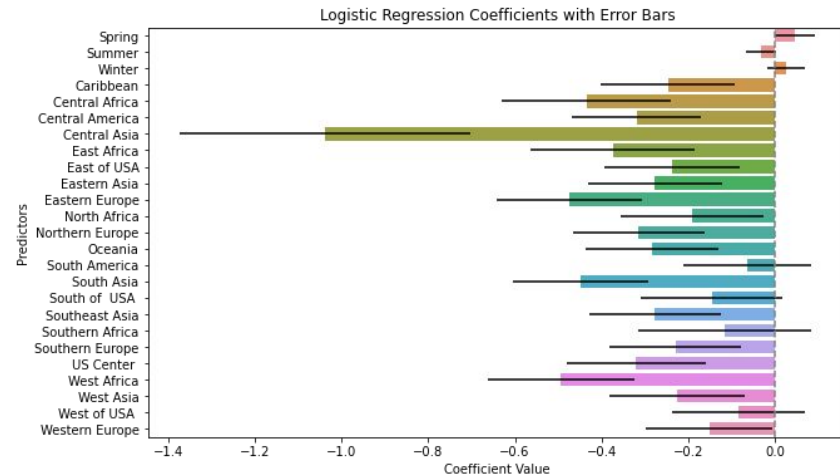
	coef	std err	t	P> t	[0.025	0.975]
const	0.0899	0.015	5.878	0.000	0.059	0.121
Spring	-0.0035	0.010	-0.337	0.738	-0.024	0.017
Summer	-0.0030	0.009	-0.325	0.747	-0.022	0.016
Winter	0.0045	0.009	0.481	0.633	-0.014	0.023
Caribbean	-0.0453	0.020	-2.211	0.033	-0.087	-0.004
Central Africa	-0.0438	0.023	-1.891	0.066	-0.091	0.003
Central America	-0.0513	0.020	-2.505	0.016	-0.093	-0.010
Central Asia	-0.0798	0.020	-3.947	0.000	-0.121	-0.039
East Africa	-0.0521	0.023	-2.246	0.030	-0.099	-0.005
East of USA	-0.0444	0.023	-1.915	0.063	-0.091	0.002
Eastern Asia	-0.0531	0.020	-2.601	0.013	-0.094	-0.012
...						



- IVs: dummies of Order Region, dummies of Season
- Approximately 45.6% of the variance in the 'Cancellation Rate' is explained by this model
- Some regions show p-values < 0.05, indicating significant differences in cancellation rates from the baseline region
- However, none of the seasons show statistically significant differences

03. Logistic Regression

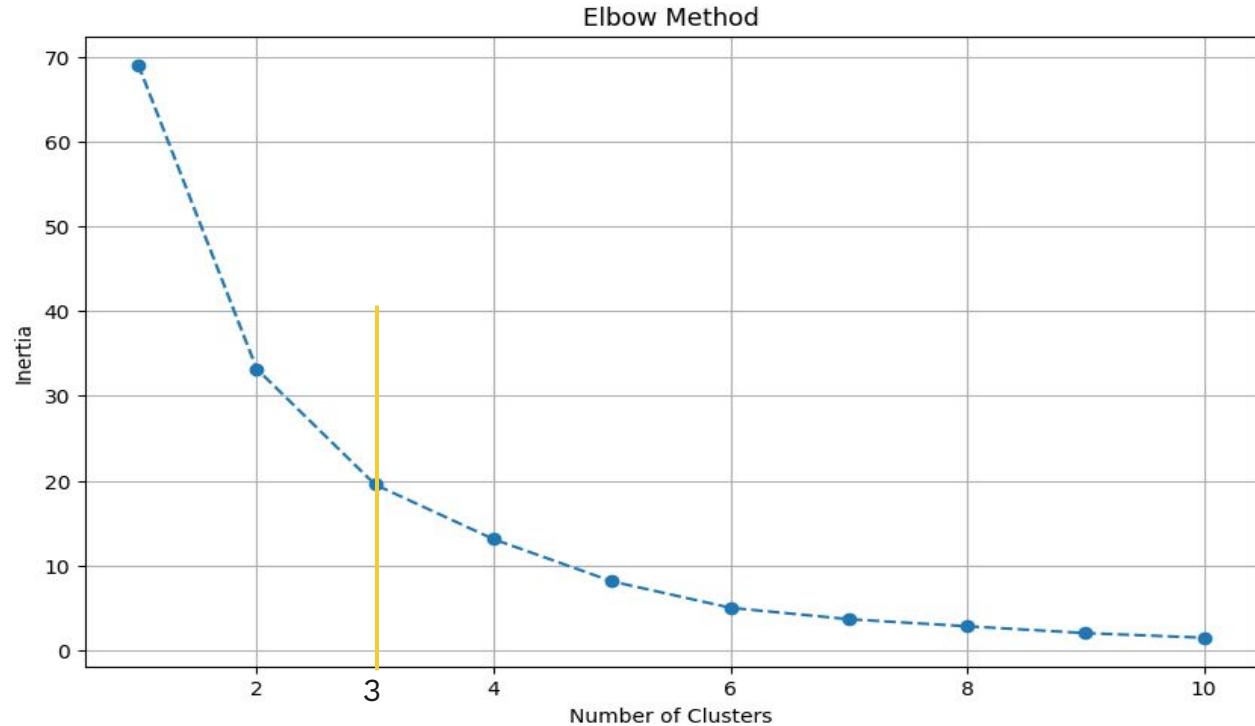
	coef	std err	z	P> z	[0.025	0.975]
const	-2.8662	0.143	-19.984	0.000	-3.147	-2.585
Spring	0.0475	0.046	1.033	0.302	-0.043	0.138
Summer	-0.0312	0.035	-0.904	0.366	-0.099	0.036
Winter	0.0256	0.043	0.599	0.549	-0.058	0.109
Caribbean	-0.2463	0.155	-1.591	0.112	-0.550	0.057
Central Africa	-0.4344	0.196	-2.217	0.027	-0.818	-0.050
Central America	-0.3187	0.148	-2.147	0.032	-0.610	-0.028
Central Asia	-1.0380	0.336	-3.086	0.002	-1.697	-0.379
East Africa	-0.3739	0.190	-1.973	0.049	-0.745	-0.002
East of USA	-0.2370	0.157	-1.510	0.131	-0.545	0.071
...						
West Asia	-0.2252	0.156	-1.444	0.149	-0.531	0.080
West of USA	-0.0840	0.154	-0.545	0.586	-0.386	0.218
Western Europe	-0.1510	0.147	-1.028	0.304	-0.439	0.137



- DV: binary value (1 = order canceled, 0 = not canceled); Same IVs
- Central Africa, Central America, Central Asia, East Africa, Eastern Europe, Northern Europe, South Asia, West Africa, and the US Center have significant effect ($p < 0.05$)
- Same as linear regression result, none of the seasons show statistically significant differences

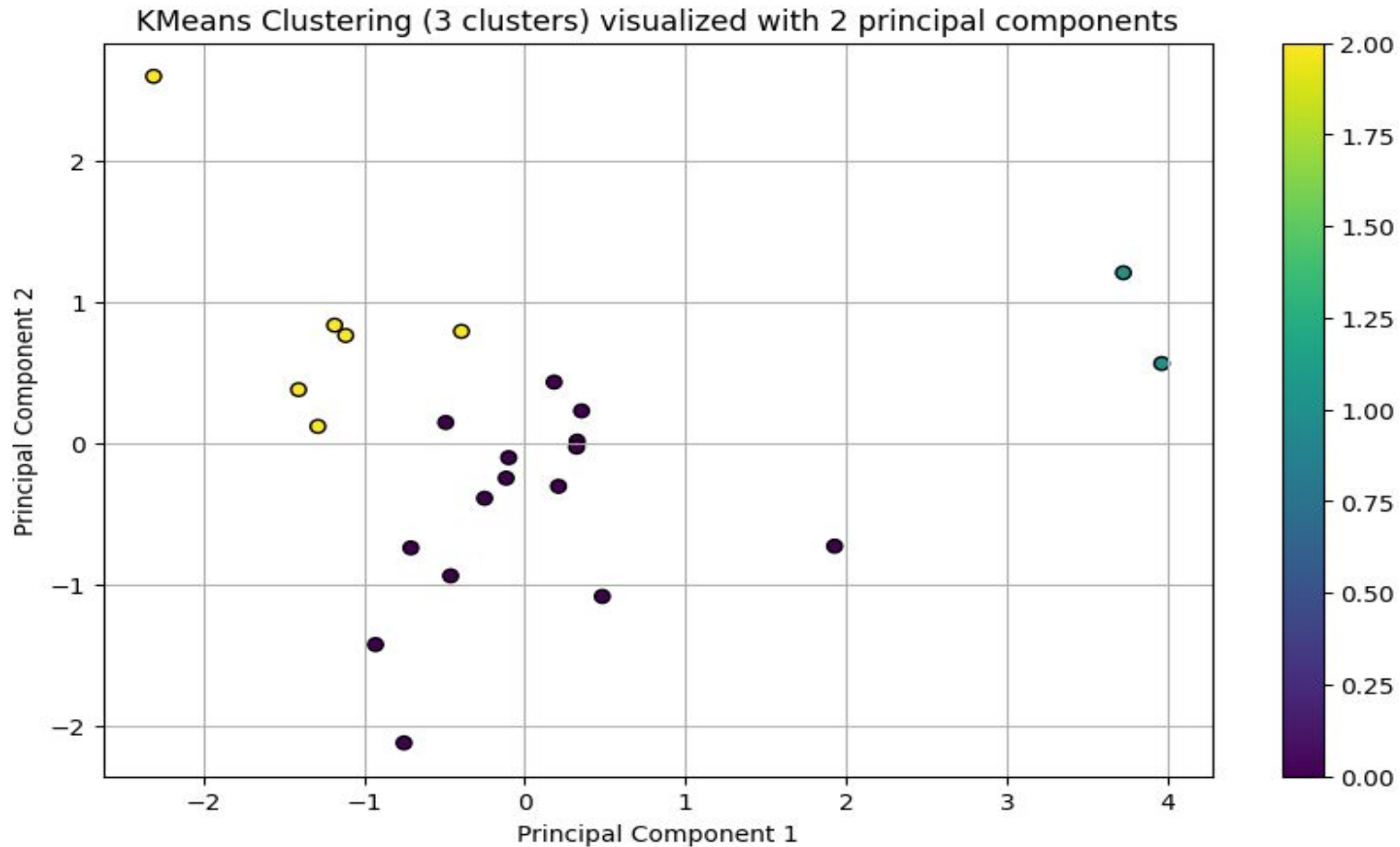
04. Clustering

K-Means Clustering was performed to divide order regions into clusters based on the cancellation ratios



Cluster	average_total_orders	average_cancellation_rate	min_Cancellation_Rate	max_Cancellation_Rate	Count
2	3238.166667	0.033094	0.019892	0.038337	6
1	27725	0.043496	0.041177	0.045815	2
0	7042.666667	0.044899	0.039216	0.054223	15

04. Clustering





Analysis Conclusion

- We failed to reject the null and conclude that the combination of seasonality and geolocation has no significant effect on the cancellation
- Order regions on the other hand have much more significance
 - This might be due to other external factors such as population density, availability of shipment modes to certain regions might be limited





Next Steps

Find Other Significant
Variables

01

Find product categories most
prone to cancellations

02



A stylized world map in light blue and white, showing continents and oceans. Four yellow location pins are placed on the map: one in North America, one in Europe, one in South America, and one in Africa. Dashed lines connect these pins, forming a network. The word "Thanks!" is centered in the middle of the map.

Thanks!

Do you have any questions?