

Clusterização de compras de e-commerce: a relação dos dados geoestatísticos com a formação de clusters

Daniele Lyra, Eduarda Rousseau e Gabriela Lea Alves

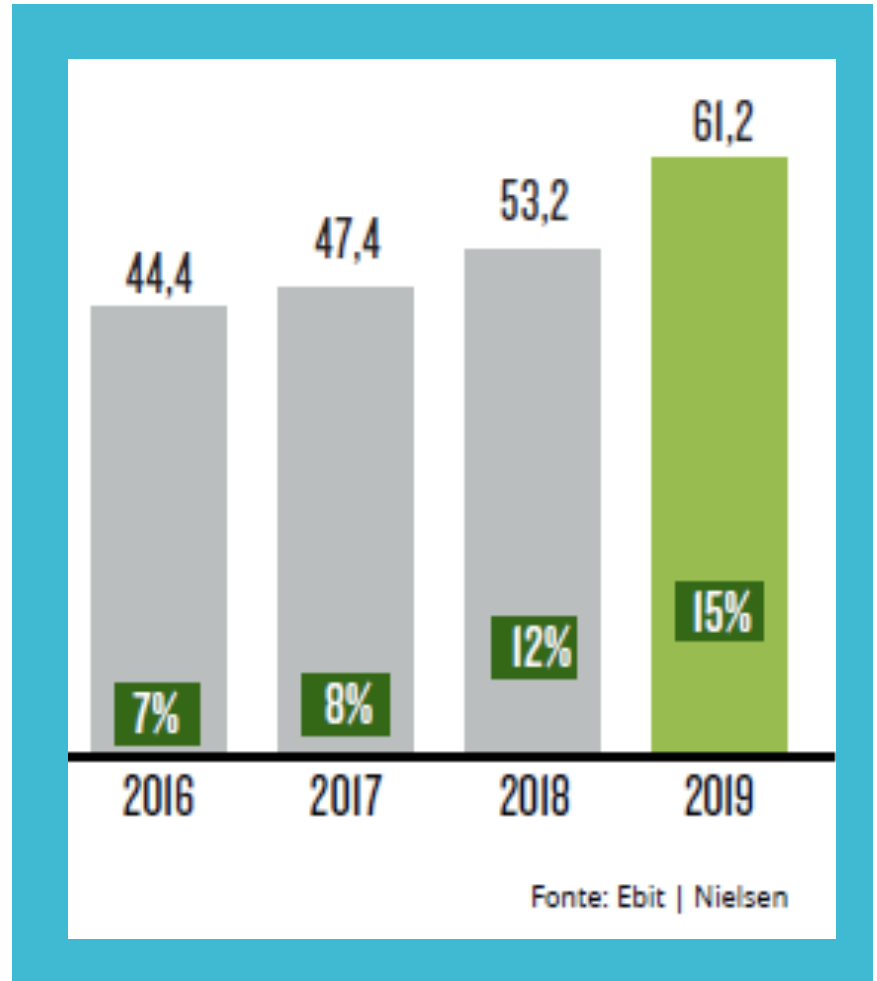
AGENDA

1. Contexto
2. Relevância
3. Objetivo
4. Dados
5. Variáveis
6. Análise Exploratória
7. Ferramentas
8. Conclusões

Contexto

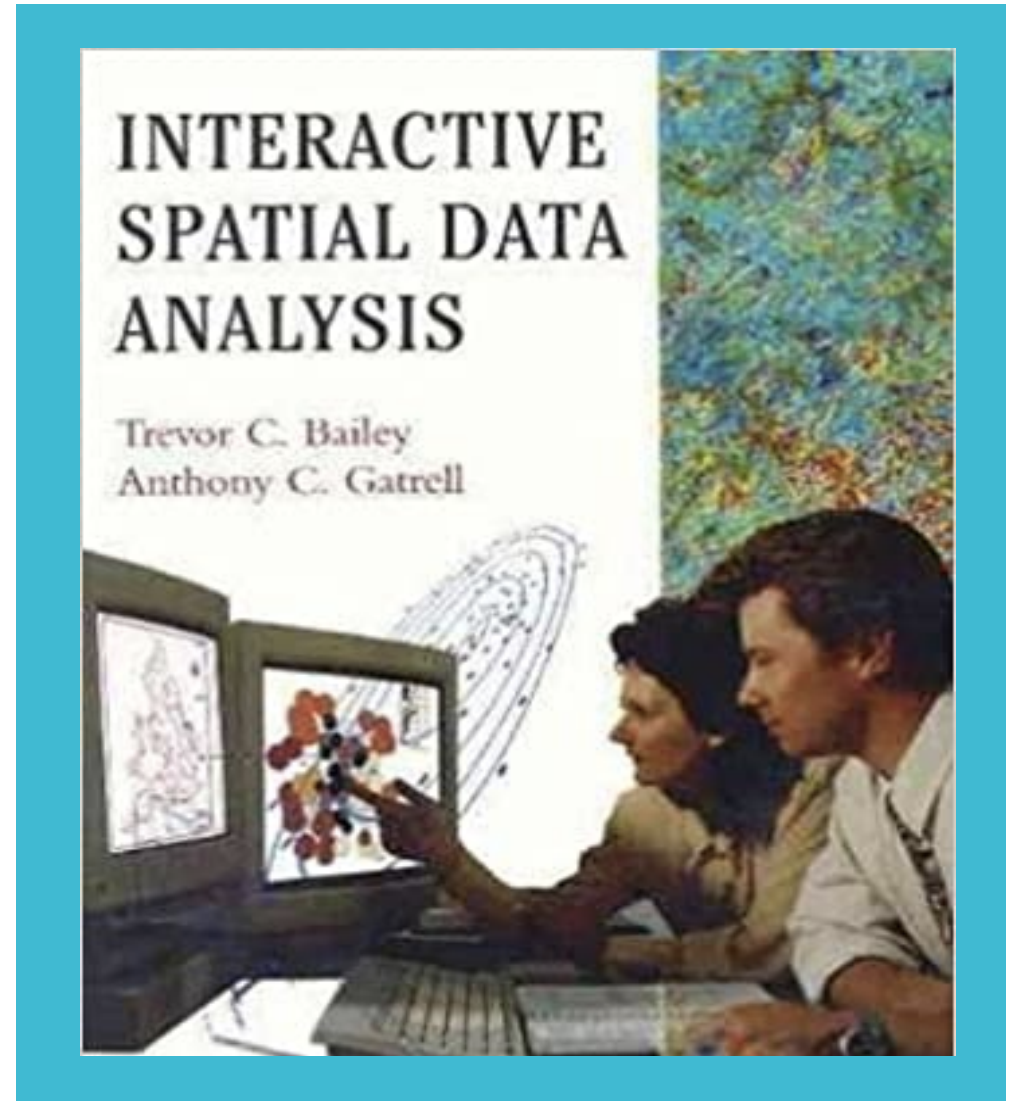
Indústria com alta taxa de crescimento

- Indústria: Varejo Online no Brasil
- Penetração de ~ 5% no varejo
- Crescimento de ~ 387% em 10 anos
- Dados de base da Olist no Kaggle



Relevância

- Identificar perfis de consumo para criação de campanhas de marketing bem segmentadas
- Oportunidade de clusterizar a base de clientes analisada



Objetivo

- Identificar se existe um perfil de consumo:
 1. Utilizando a base original com técnicas de clusterização
 2. Identificar padrões de consumo correlacionado a bancos de dados públicos

Dados

Base de Dados

- Base: ~ 18 mil pedidos
- Período: de 2016 a 2018
- Geografia: cidade de São Paulo sobre a ótica de subdistritos
- Base de Dados do IBGE: Censo Geográfico de SP de 2010

Variáveis

Base Olist 2016 - 2018

- Quantidade de Pedidos (unidade)
- Valor do Pedido (R\$)
- Desvio padrão dos pedidos na localidade
- Valor do Frete (R\$)

Censo IBGE 2010

- Renda da população (R\$)
- Moradores por residência

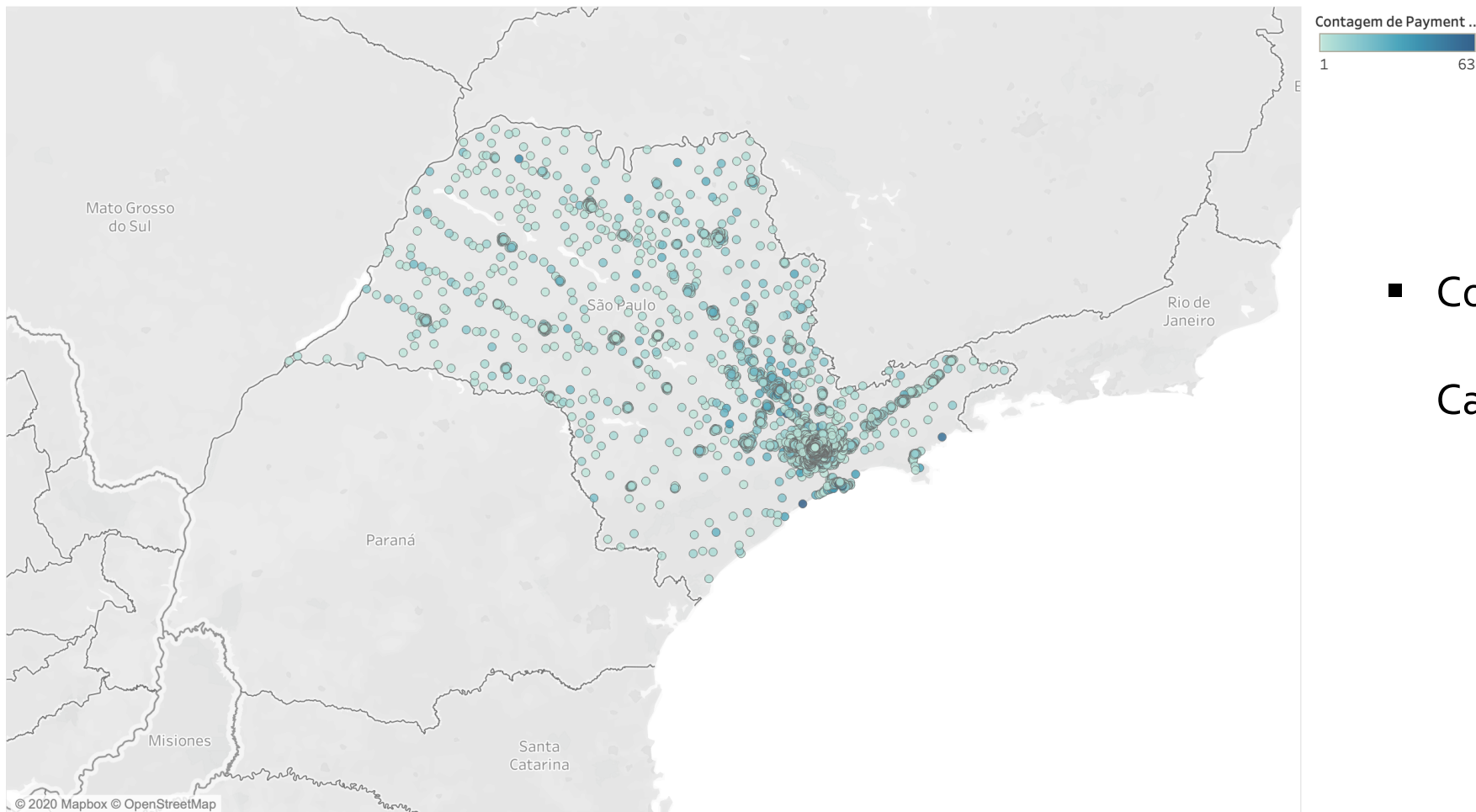
Ferramentas



Análise Exploratória

Distribuição de Pedidos no Estado de SP

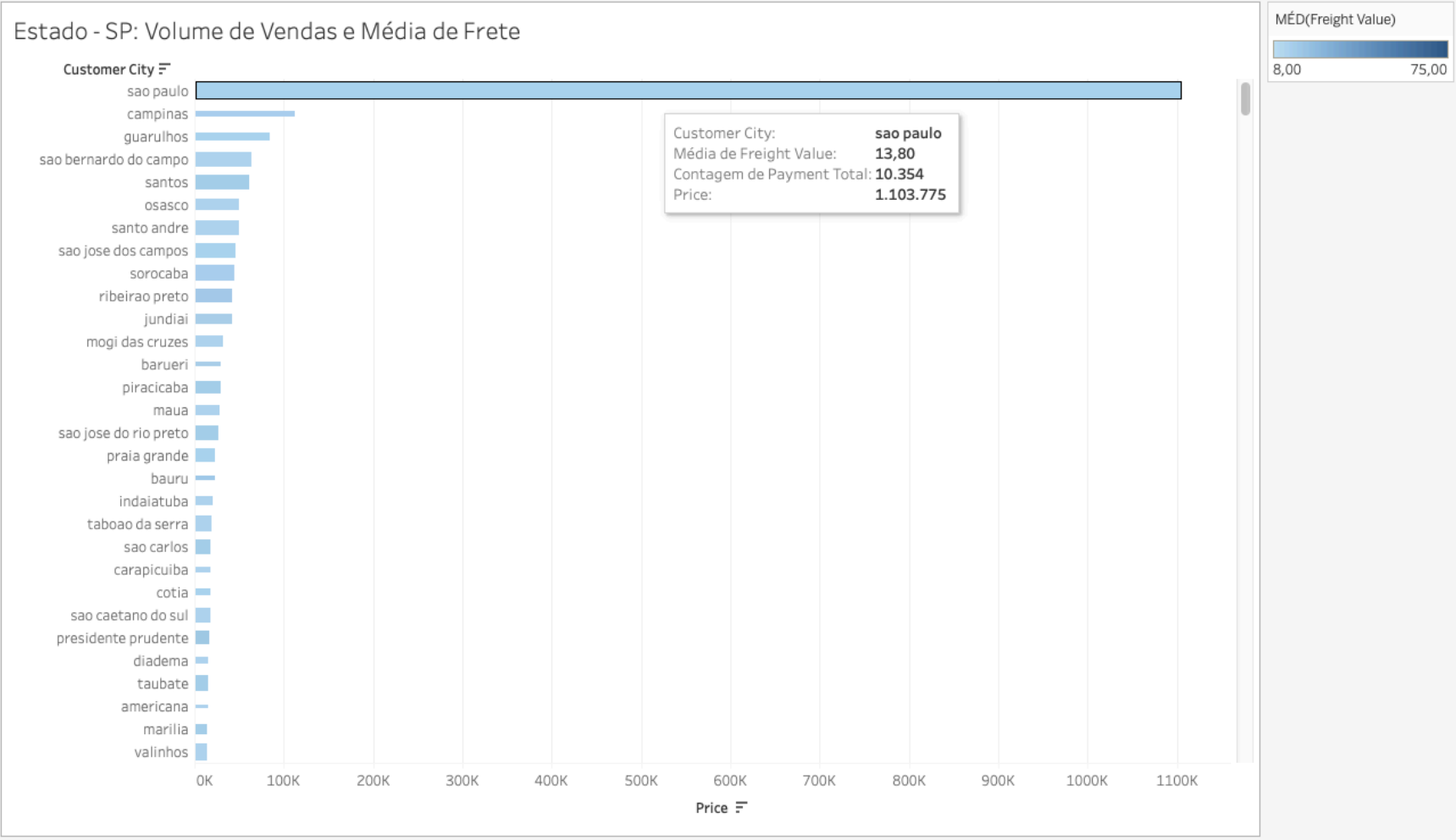
Estado



- Concentração de pedidos na Capital do Estado

Mapa baseado em Longitude (gerada) e Latitude (gerada). A cor mostra contagem de Payment Total. Detalhes são mostrados para Customer Zip Code Prefix. Os dados estão filtrados em Customer State, que mantém SP. A exibição está filtrada em Customer Zip Code Prefix, que exclui 8 membros.

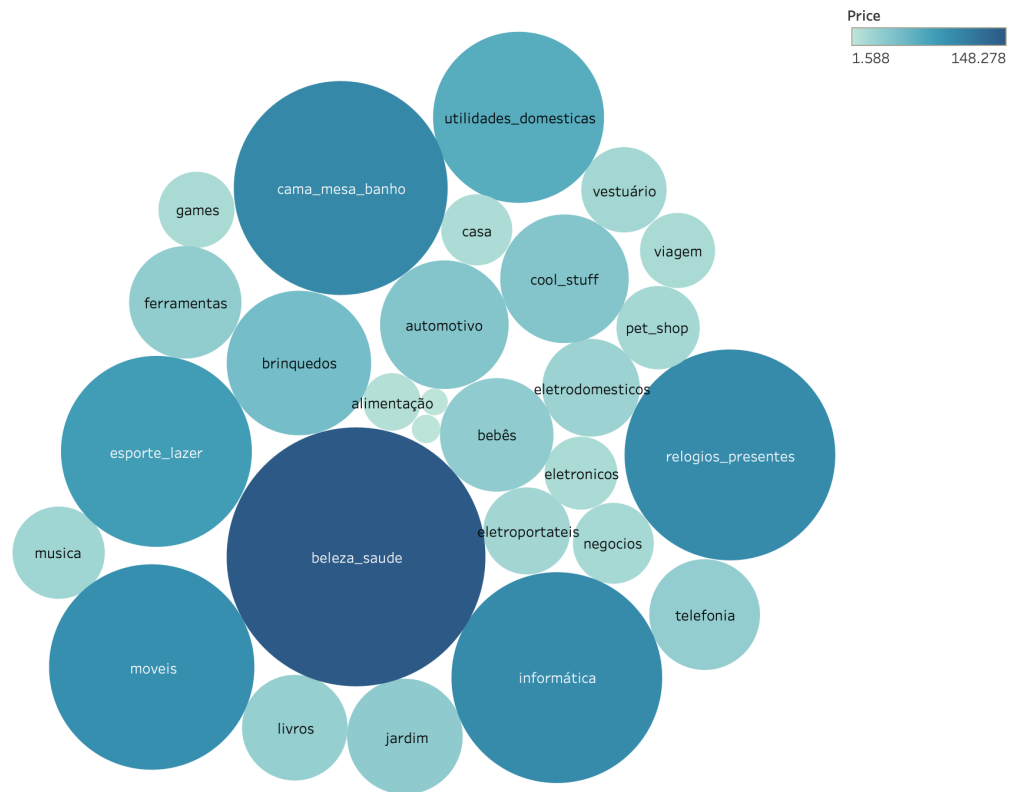
Análise de Métricas no Estado de SP



- Relevância da Capital para a análise dos dados

Foco em Cama Mesa e Banho e Beleza

Capital SP: Categorias mais Vendidas

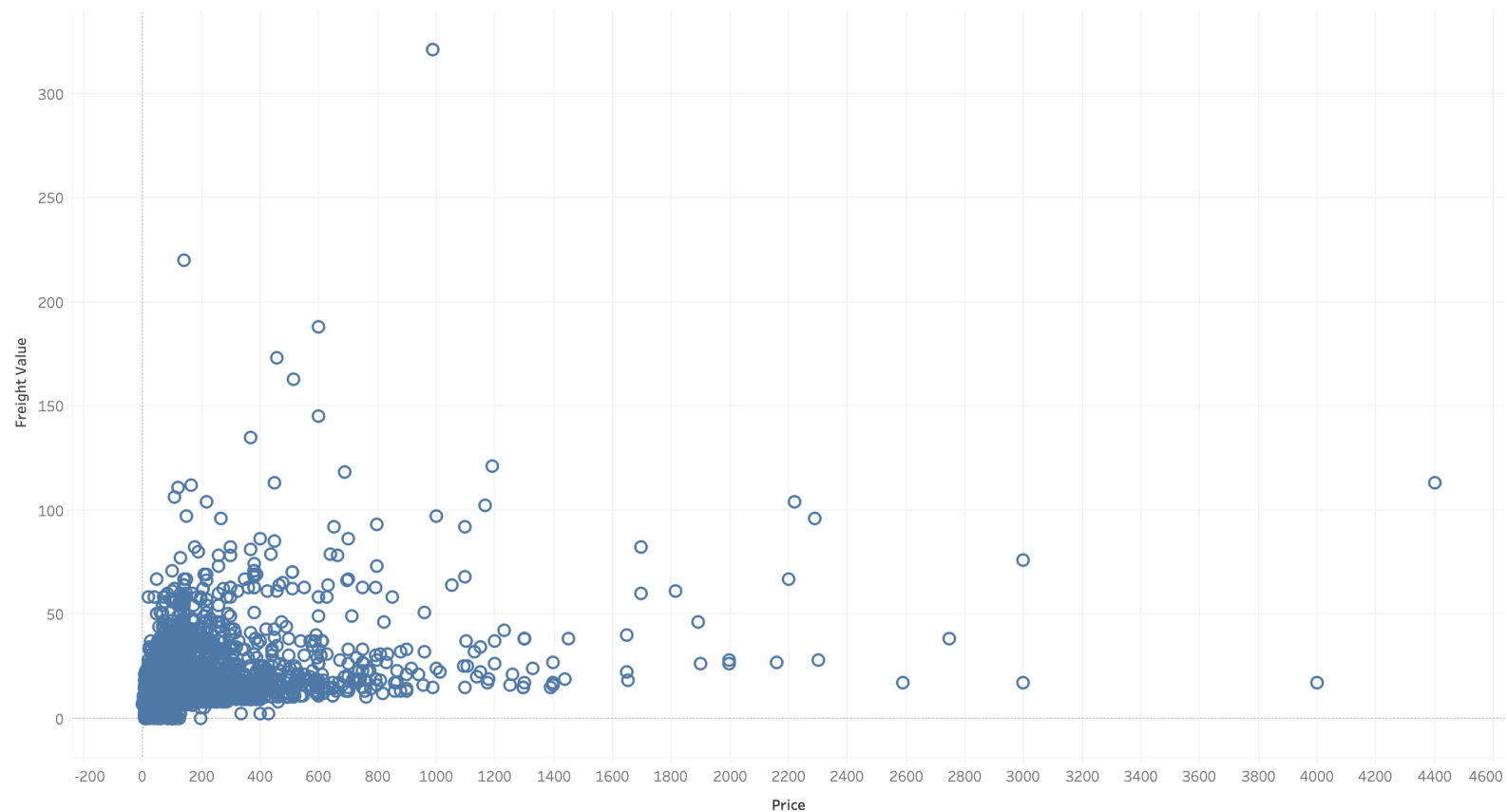


Product Category Name (grupo). A cor mostra soma de Price. O tamanho mostra soma de Price. As marcas são rotuladas por Product Category Name (grupo). Os dados estão filtrados em Customer City, que mantém sao paulo.

- Categorias relevantes para a distribuição

Distribuição de Vendas

Capital SP: Valor Produto x Valor Frete



Price vs. Freight Value. Detalhes são mostrados para Customer Zip Code Prefix e Product Category Name. Os dados estão filtrados em Customer City, que mantém sao paulo.

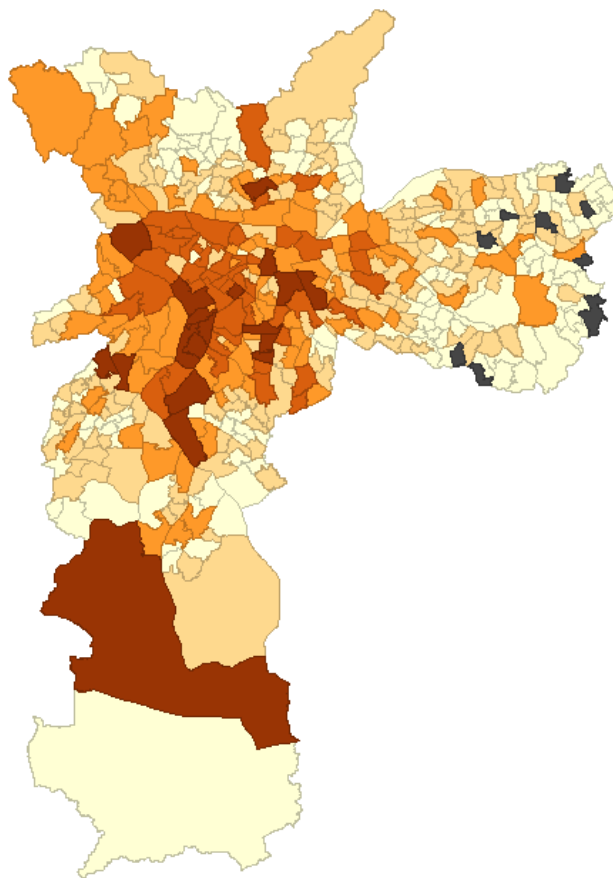
- Relevância da
Capital para a
análise dos dados

Conclusões

Método Natural Breaks de Pedidos

Natural Breaks: customer_u

< 22	(144)
[22, 42)	(149)
[42, 68)	(87)
[68, 108)	(48)
>= 108	(19)
undefined	(9)

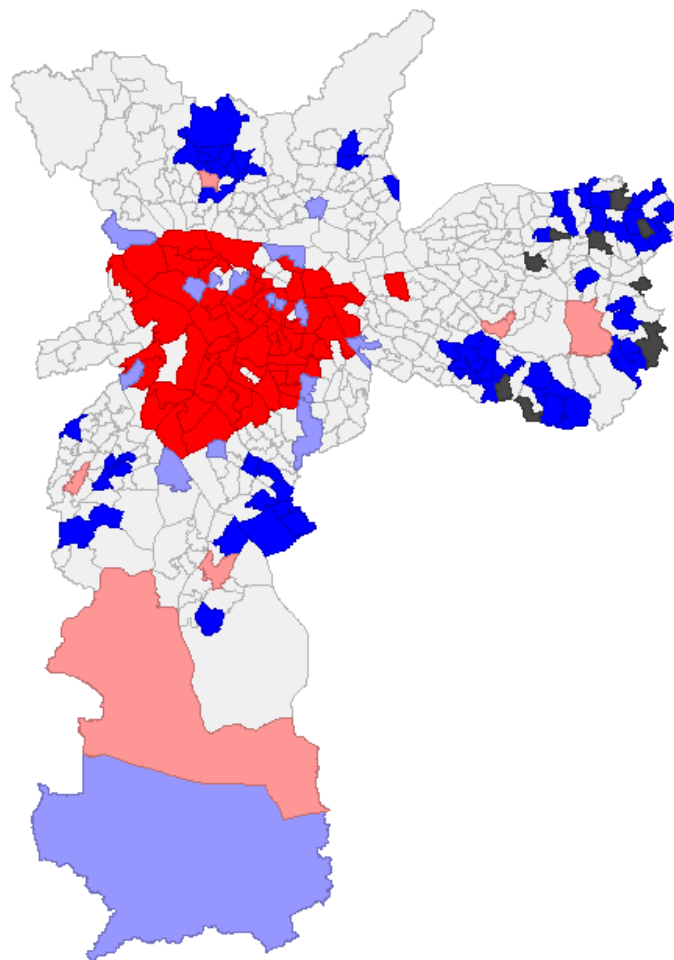


- Concentração mais relevante no Centro-oeste
- Ponto estratégico na Zonal Sul

Matriz de Vizinhanças da Quantidade de Pedidos

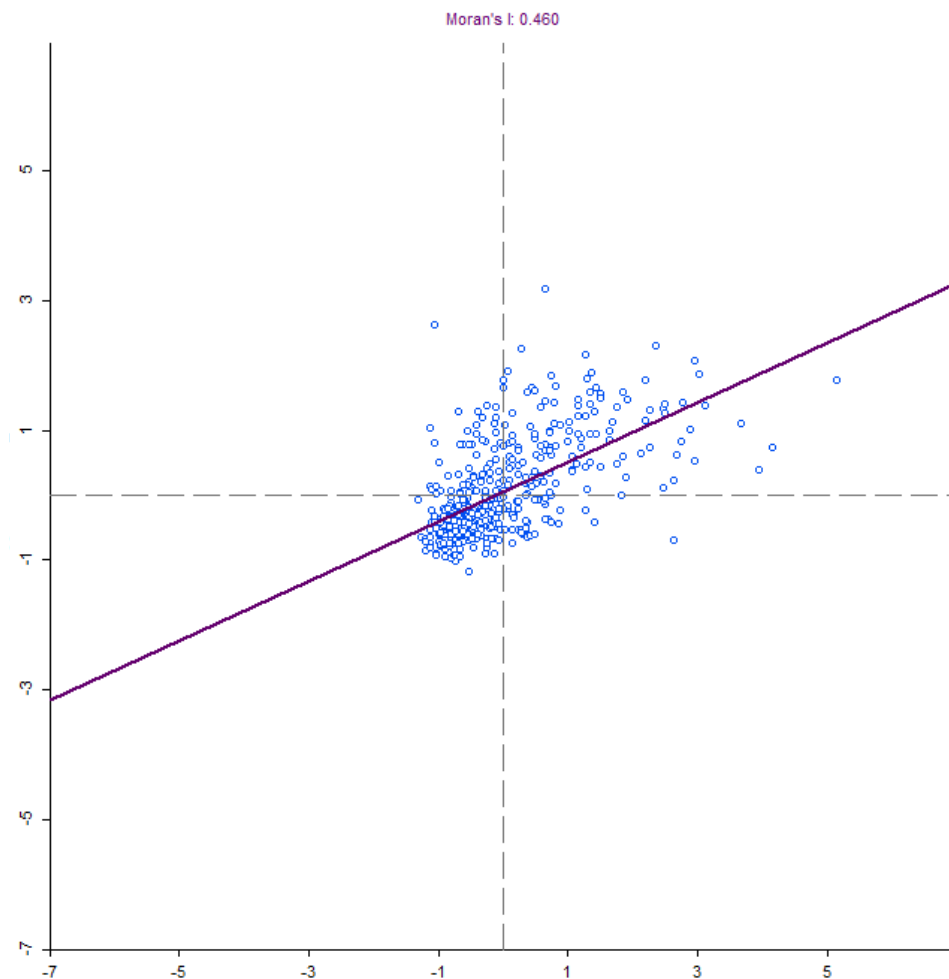
LISA Cluster Map: geogeos

Not Significant (296)
High-High (69)
Low-Low (60)
Low-High (16)
High-Low (6)
Undefined (9)



- Centro de São Paulo tem alto volume de pedidos
- Crescente de volume na Zona Sul próxima ao Centro

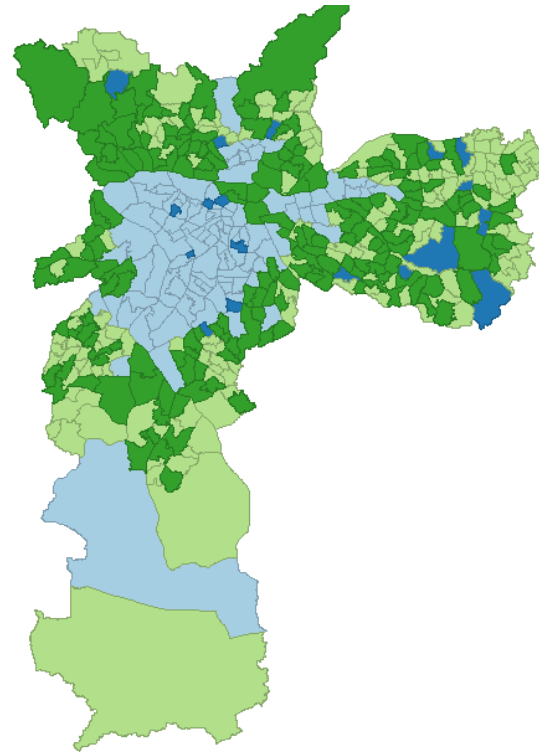
Índice de Moran



- Correlação espacial significativa da variável de quantidade de pedidos (und) com a Renda da população

Clusterização

1 (102)
2 (18)
3 (158)
4 (178)



Clusterização dos Clientes por Fidelização

Azul claro: VIP

Verde claro: Insiders

Azul escuro: Visitors

Verde escuro: Unprobable

Conclusão

Decisão da Diretoria de Marketing

- Clusterização foco clientes VIP e Insiders
- Campanha para as categorias Beleza e Saúde e Cama Mesa e Banho
- Promoção Frete Grátis Zona Sul