

Olá!

Este documento foi criado com o intuito de documentar os passos feitos e as conclusões chegadas de acordo com a análise dos dados obtidos por processo de mentoria.

As informações e conclusões as quais o presente documento chegaram não possuem valor mercadológico ou qualquer intuito comercial, mas tão somente acadêmico e para fins de demonstração de skills de Análise de Dados para Negócios com fito de gerar portfólio.

As informações apresentadas serão feitas na abordagem *top-down*, ou seja, primeiro serão mostradas as conclusões inferidas e após mostrar-se-á todo o projeto para chegar a tal conclusão, desde linhas de código até *prints* dos resultados.

O processo de ETL (Extração, Transformação e Limpeza) dos Dados e de criação de tabelas no *database* não será parte deste documento mas será internalizado em artigo em site próprio.

As ferramentas usadas para o presente projeto foram:

Oracle MySQL 8.0

Libre Office Calc

Libre Office Writer

Foxit PDF

Microsoft Windows 10

Análise Exploratória de Dados

Inicial

1. Quantas linhas e colunas existem na base?
Linhas: 48783 (após processo de ETL onde algumas informações inconsistentes precisaram ser excluídas)
Colunas: 29 (após processo de ETL onde algumas informações inconsistentes precisaram ser excluídas)
2. É possível classificar as colunas em entidades?
Sim, não é necessário realizar joins com outras tabelas.
3. Do que se trata a base de dados?
Base de dados com informações de pedidos de entregas.
4. Quantos pedidos existem na base?
48.783
5. Quantos usuários existem na base?
17.280
6. Verifique se todos os pedidos são únicos.
Sim, ao selecionar os pedidos distintos retorna o mesmo valor da quantidade da questão 4.
7. Qual o pedido mais recente e o mais antigo na base?
Mais recente: 31/12/2019; Mais antigo: 01/06/2019
8. Quais foram os valores mínimo, máximo e médio gasto por todos os clientes durante os meses disponíveis na base?
Mínimo: 13.05; Máximo: 667.70; Médio: 55.35
9. Qual o dispositivo que faz mais pedidos?
Mobile
10. Qual sistema operacional é mais usado para os pedidos?
Android
11. Qual o estado que faz mais pedidos?
São Paulo, com 18.705 pedidos.
12. No estado mais pedido, qual a cidade que faz mais pedidos?
São Paulo – Capital, com 8.103 pedidos
13. Qual a quantidade e percentual de pedidos por turno?
Maior: Jantar FDS com 32,72%; Menor: Café da Manhã FDS com 0,09%
14. Qual a quantidade e percentual de pedidos por tipo?
Lanches com 29,48%; Comida Brasileira com 18,31%; Pizza com 16,91%
15. Qual o percentual de clientes que são novos na base?
1,59%

Intermediário

16. Qual a quantidade de itens pedidos, em média?
1,6
17. Qual a média de valores dos pedidos?
R\$55,35
18. Qual o percentual de pedidos que são agendados?
1,40%
19. Qual o tipo de pedido preferido por turno?
Comida Saudável: almoço
Doces & Bolos: café da manhã
Lanches: café da manhã fds, jantar, madrugada
Comida Italiana: lanche durante a semana
Pizza: lanche fds
20. Em qual turno dos dias se tem o maior ticket médio? (média de valor de pedidos)
jantar fds com uma média de 61,85 seguida pelo almoço fds com 61,19
21. Em qual tipo de pedidos se tem o maior ticket médio? (média de valor de pedidos)
pizza, pela proporção de restaurantes desse tipo, com o valor médio de 63,57
22. Para clientes sensíveis a cupons, qual tem o maior ticket médio? (média de valor de pedidos)
o sensibilidade baixa, comprando a média de valor de 65,24
23. Para clientes sensíveis a cupons, qual o principal tipo de pedidos?
Lanches, Comida Brasileira, Pizza e Comida Japonesa

Avançado

24. Como analista da empresa, os restaurantes de Sorocaba que vendem lanches querem saber qual o melhor turno para ofertar um cupom, com base na sensibilidade dos clientes e do ticket médio. considerando SQL a resposta seria 'jantar fds' (dash em breve)

CONSULTAS E RESULTADOS

-- 1) Quantas linhas e colunas existem na base?

```
SELECT COUNT(id_pedido) AS qtd_linhas_tb FROM base_pedidos; -- linhas
```

```
SELECT COUNT(*) AS qtd_colunas FROM information_schema.COLUMNS  
WHERE TABLE_NAME = 'base_pedidos'; -- colunas
```

Result Grid	
	qtd_linhas_tb
▶	48783

Result Grid	
	qtd_colunas
▶	29

-- 4) Quantos pedidos existem na base?

```
SELECT COUNT(id_pedido) AS qtd_pedidos  
FROM base_pedidos;
```

Result Grid	
	qtd_pedidos
▶	48783

-- 5) Quantos usuários existem na base?

```
SELECT COUNT(DISTINCT(id_cliente)) AS qtd_clientes  
FROM base_pedidos;
```

Result Grid	
	qtd_clientes
▶	17280

-- 7) Qual o pedido mais recente e o mais antigo da base?

```
SELECT id_pedido, CAST(horario_local AS DATE) AS mais_antigo  
FROM base_pedidos  
GROUP BY horario_local  
ORDER BY horario_local ASC  
LIMIT 3; -- mais antigo
```

```
SELECT id_pedido, CAST(horario_local AS DATE) AS mais_antigo  
FROM base_pedidos  
GROUP BY horario_local  
ORDER BY horario_local DESC  
LIMIT 3; -- mais recente
```

	id_pedido	mais_recente
▶	71b5ce4c-6ca5-4f96-9d85-ed6a2cdac4fe	2019-12-31
	66476a01-c1f1-488b-9902-7910cd86a985	2019-12-31
	1f10246d-8c0b-44ae-b4a0-6019566ca6bc	2019-12-31

	id_pedido	mais_antigo
▶	b715c516-8e5d-40ad-be14-1696bd868cd7	2019-06-01
	9ad086f0-6faa-4c60-ba26-fd530c7d1e9b	2019-06-01
	0d1c88c3-4aaa-4dc8-bbc6-5648c80b6ff3	2019-06-01

/*8) Quais foram os valores mínimo, máximo e médio gasto por todos os clientes durante os meses disponíveis na base? */

```
SELECT MIN(total_pedido) AS minimo,  
MAX(total_pedido) AS maximo,  
ROUND(AVG(total_pedido),2) AS medio  
FROM base_pedidos;
```

Result Grid			
	minimo	maximo	medio
▶	13.05	667.7	55.35

-- 9)Qual o dispositivo que faz mais pedidos?

```
SELECT tipo_dispositivo, COUNT(tipo_dispositivo) AS tipo_distinto  
FROM base_pedidos  
GROUP BY tipo_dispositivo  
ORDER BY COUNT(tipo_dispositivo) DESC;
```

Result Grid		
	tipo_dispositivo	tipo_distinto
▶	MOBILE	47544
	SITE	1239

-- 10) Qual sistema operacional é mais usado para os pedidos?

```
SELECT sistema_operacional, COUNT(sistema_operacional) AS sistema  
FROM base_pedidos  
GROUP BY sistema_operacional  
ORDER BY COUNT(sistema_operacional) DESC;
```

Result Grid		
	sistema_operacional	sistema
▶	ANDROID	27340
	IOS	20481
	DESKTOP	962

```
-- 11) Qual o estado que faz mais pedidos?
SELECT cliente_estado, COUNT(cliente_estado) AS estados
FROM base_pedidos
GROUP BY cliente_estado
ORDER BY COUNT(cliente_estado) DESC
LIMIT 10;
```



Result Grid	Filter Rows:
cliente_estado	estados
SP	18705
RJ	6471
MG	4204
PR	3525
DF	1939
CE	1896
RS	1806
SC	1676
PE	1468
ES	1069

```
-- 12) No estado mais pedido, qual a cidade que faz mais pedidos?
SELECT cliente_cidade, COUNT(cliente_cidade) AS cidades
FROM base_pedidos
WHERE cliente_estado = 'SP'
GROUP BY cliente_cidade
ORDER BY COUNT(cliente_cidade) DESC
LIMIT 10;
```

Result Grid	Filter Rows:
cliente_cidade	cidades
SAO PAULO	8103
CAMPINAS	1202
RIBEIRAO PRETO	719
SAO JOSE DOS CAMPOS	584
GUARULHOS	555
SANTO ANDRE	499
SANTOS	436
SAO BERNARDO DO CAMPO	394
OSASCO	368
SOROCABA	335



-- 13. Qual a quantidade e percentual de pedidos por turno?

```
SELECT turno_pedido, ROUND((COUNT(turno_pedido)*100/tb1.turno_do_pedido),2) AS percent
FROM base_pedidos, (SELECT COUNT(turno_pedido) AS turno_do_pedido FROM base_pedidos) AS tb1
GROUP BY turno_pedido
ORDER BY 2 DESC;
-- 100*medida/totaldedadosbase
```

Result Grid   Filter Rows: <input type="text"/>		
	turno_pedido	percent
▶	jantar fds	32.72
	jantar durante a semana	27.95
	almoço durante a semana	21.87
	almoço fds	9.06
	lanche durante a semana	3.57
	lanche fds	2.04
	madrugada fds	1.43
	madrugada durante a semana	1.00
	café da manhã durante a semana	0.27
	café da manhã fds	0.09



-- 14. Qual a quantidade e percentual de pedidos por tipo (Comida brasileira, lanches, etc)?

```
SELECT comercio_tipo, ROUND((COUNT(comercio_tipo)*100/tb1.tipo_do_comercio),2) AS percent
FROM base_pedidos, (SELECT COUNT(comercio_tipo) AS tipo_do_comercio FROM base_pedidos) AS tb1
GROUP BY comercio_tipo
ORDER BY 2 DESC
LIMIT 10;
```

Result Grid   Filter Rows: <input type="text"/>		
	comercio_tipo	percent
▶	Lanches	29.48
	Comida Brasileira	18.31
	Pizza	16.91
	Comida Japonesa	7.07
	Comida Saudável	3.78
	Alçaça	3.28
	Comida Chinesa	2.49
	Comida Italiana	2.40
	Marmita	1.88
	Comida Variada	1.84

-- 15. Qual o percentual de clientes que são novos na base?

```
SELECT cliente_status_ultimo_mes, ROUND((COUNT(cliente_status_ultimo_mes)*100/tb1.cliente_mes),2) AS percent
FROM base_pedidos, (SELECT COUNT(cliente_status_ultimo_mes) AS cliente_mes FROM base_pedidos) AS tb1
WHERE cliente_status_ultimo_mes = 'New'
GROUP BY cliente_status_ultimo_mes;
```

Result Grid   Filter Rows: <input type="text"/>		
	cliente_status_ultimo_mes	percent
▶	New	1.59

```
-- Intermediário
-- 16. Qual a quantidade de itens pedidos, em média?
SELECT * FROM base_pedidos
LIMIT 50;
```

```
SELECT ROUND(AVG(qtd_itens),1) AS media_pedidos
FROM base_pedidos;
```

Result Grid	
	media_pedidos
▶	1.6

```
-- 17. Qual a média de valores dos pedidos?
SELECT ROUND(AVG(total_pedido),2) AS media_valor_pedido
FROM base_pedidos;
```

Result Grid	
	media_valor_pedido
▶	55.35

```
-- 18. Qual o percentual de pedidos que são agendados?
SELECT agendada, ROUND((COUNT(agendada)*100/tb1.pedido_agendado),2) AS percent
FROM base_pedidos, (SELECT COUNT(agendada) AS pedido_agendado FROM base_pedidos) AS tb1
GROUP BY agendada
ORDER BY 2 DESC
LIMIT 10;
```

Result Grid		
	agendada	percent
▶	False	98.60
	True	1.40

```
-- 19. Qual o tipo de pedido preferido por turno?
SELECT comercio_tipo, turno_pedido, COUNT(comercio_tipo) AS tipo_de_comida
FROM base_pedidos
GROUP BY turno_pedido
ORDER BY turno_pedido;
```

Result Grid			
	comercio_tipo	turno_pedido	tipo_de_comida
▶	Comida Saudável	almoço durante a semana	10670
	Comida Saudável	almoço fds	4420
	Doces & Bolos	café da manhã durante a semana	131
	Lanches	café da manhã fds	42
	Lanches	jantar durante a semana	13636
	Lanches	jantar fds	15963
	Comida Italiana	lanche durante a semana	1740
	Pizza	lanche fds	994
	Lanches	madrugada durante a semana	487
	Lanches	madrugada fds	700

```
-- 20. Em qual turno dos dias se tem o maior ticket médio? (média de valor de pedidos)
-- R$55,35

SELECT turno_pedido, ROUND(AVG(total_pedido),2) AS media_valor_pedido
FROM base_pedidos
GROUP BY turno_pedido
HAVING media_valor_pedido >= 55
ORDER BY media_valor_pedido DESC;
```

Result Grid			Filter Rows:	
	turno_pedido	media_valor_pedido		
▶	jantar fds	61.85		
	almoço fds	61.19		
	lanche fds	58.16		
	jantar durante a semana	56.82		

```
-- 21. Em qual tipo de pedidos se tem o maior ticket médio? (média de valor de pedidos)

SELECT comercio_tipo,
COUNT(comercio_tipo) AS qtd_de_tipo,
ROUND(AVG(total_pedido),2) AS media_valor_pedido
FROM base_pedidos
GROUP BY comercio_tipo
HAVING media_valor_pedido >= 55
ORDER BY qtd_de_tipo DESC
LIMIT 5;
```

Result Grid				Filter Rows:		Exp
	comercio_tipo	qtd_de_tipo	media_valor_pedido			
▶	Pizza	8247	63.57			
	Comida Japonesa	3448	72.11			
	Comida Chinesa	1214	64.37			
	Comida Italiana	1172	67.51			
	Comida Variada	899	56.56			

⊕ /*/22. Para clientes sensíveis a cupom, qual tem o maior ticket médio?

- SELECT cliente_sensivel_cupom AS sensibilidade_cupom,
COUNT(cliente_sensivel_cupom) AS qtd,
ROUND(AVG(total_pedido),2) AS média_de_pedido
FROM base_pedidos
WHERE cliente_sensivel_cupom IN ('media','alta')
GROUP BY cliente_sensivel_cupom
ORDER BY 3 DESC;

Result Grid				Filter Rows:	
	sensibilidade_cupom	qtd	média_de_pedido		
▶	Media	15157	62.39		
	Alta	25125	48.29		

```
-- 23. Para clientes sensíveis a cupons, qual o principal tipo de pedidos?
SELECT comercio_tipo, COUNT(comercio_tipo) AS qtd_tipo, cliente_sensivel_cupom
FROM base_pedidos
WHERE cliente_sensivel_cupom IN ('media','alta')
GROUP BY comercio_tipo
ORDER BY 2 DESC;
```

	comercio_tipo	qtd_tipo	cliente_sensivel_cupom
▶	Lanches	11846	Alta
	Comida Brasileira	7647	Alta
	Pizza	6715	Alta
	Comida Japonesa	2857	Media
	Comida Saudável	1442	Alta
	Alça	1331	Alta
	Comida Chinesa	954	Media
	Comida Italiana	941	Alta
	Marmita	797	Alta
	Comida Variada	725	Alta

-- Avançado

/* 24. Como analista da empresa, os restaurantes de Sorocaba que vendem lanches querem saber qual o melhor turno para ofertar um cupom, com base na sensibilidade dos clientes e do ticket médio. (Pode embasar a resposta com tabelas, gráficos, etc)*/

```
SELECT turno_pedido,
       COUNT(id_pedido) AS qtd_pedidos,
       ROUND(AVG(total_pedido),2) AS media_pedido,
       cliente_sensivel_cupom AS sensibilidade
FROM base_pedidos
WHERE comercio_tipo = 'Lanches'
      AND comercio_cidade = 'sorocaba'
      AND cliente_sensivel_cupom IN ('media','alta')
GROUP BY turno_pedido
      HAVING media_pedido > 55
ORDER BY qtd_pedidos DESC;
```

	turno_pedido	qtd_pedidos	media_pedido	sensibilidade
▶	jantar fds	36	55.09	Media
	lanche durante a semana	4	57.72	Alta
	lanche fds	1	56.8	Alta