

Análise dos dados da Discorama

Analista/Engenheiro de Analytics: João Vitor Pinheiro Ocker

Quais são as perguntas de negócio relevantes para esse projeto de analytics?

1. Como a Discorama pode elevar as locações e as vendas de DVD? Existe alguma categoria de filmes e/ou atores com maior probabilidade de alcançar esses objetivos?
2. O que pode ser feito para reduzir a taxa de perda do inventário?
3. Quais os valores totais de locações, vendas e ticket médio? E como estão divididos esses valores entre os diferentes níveis de granularidade, como categorias do filme, filme e atores?
4. Como está distribuído as locações, vendas e ticket médio entre os meses do ano? Algum mês é mais expressivo?
5. Qual loja apresenta o maior volume de locações, de vendas e de ticket médio?
6. Qual o valor dos indicadores para cada cliente, como locações, vendas etc.?

Quais serão os indicadores ou visualizações utilizadas para responder às perguntas de negócio citadas anteriormente?

Indicadores: Vendas, Ticket médio, locações, Taxa de perda de inventário, filmes não devolvidos;

Visualizações: Gráfico de barras, Gráfico de linhas, Gráfico de colunas, árvore hierárquica, cartões;

A partir dos indicadores, das visualizações, dos filtros e com base em diferentes granularidades, é possível encontrar as respostas para os questionamentos acima no arquivo .pbix disponibilizado.

Apresentar pelo menos duas análises de negócio para cada uma das categorias:

Uma análise diagnóstica;

- Entender a performance das categorias de filme para observar quais filmes são mais populares entre os clientes, com criações de clusters de clientes com base em suas preferências, além de como a localização geográfica afeta suas escolhas.

- Entender a *churn rate* com base no histórico de aluguéis dos clientes, nos padrões de filmes alugados, na sua atividade na conta da locadora. Tais medidas podem evitar a perda de clientes.

Uma análise descritiva:

- Análise do total de vendas em diferentes meses nos anos disponibilizados na amostra de dados;

- Análise do total de locações e ticket médio por diferentes clientes;

- Análise de qual gênero/categoria de filmes obtiveram maior número de locações e faturamento;
- Análise de quais atores são campeões de vendas nos dados.

Uma análise preditiva;

- Criar um modelo de previsão de vendas com técnicas de séries temporais, com base nos dados de pagamentos e locações dos clientes;
- Criar um algoritmo como uma árvore da decisão que classifique os clientes em cluster para avaliar quais categorias de filme eles teriam maior probabilidade de aluguel.
- Criar um algoritmo avançado com técnicas de *deep learning* que podem ajudar nas recomendações de filmes novos a serem adquiridos, com base nos padrões de uso dos clientes e suas preferências.

Uma análise prescritiva.

- Preços personalizados e promoções por categoria de filmes, com base no comportamento histórico do cliente e tendências de mercado para elevação das vendas e locações.
- Novas aquisições de filmes com base nos custos de licenciamento de novos produtos, comportamento e preferências de consumo dos clientes, além dos resultados dos algoritmos acima.
- O tempo ideal para o lançamento de filmes com base em localidades geográficas, padrões dos clientes e até mesmo eventos específicos.

Quais os processos de transformação e tratamento de dados foram aplicados?

- Criação de um repositório no Github, e posteriormente o clone na máquina local com Git bash;
- Verificação de forma preliminar das tabelas fatos e dimensões no Excel, com a utilização de formulas como vlookup, index e match para a mesclar as tabelas e funções de agregações como count, average e sum para obter totais específicos;
- Criação do modelo de dados no Power BI entre as tabelas fato e dimensões para o melhor entendimento dos dados, com a mudança dos tipos de dados para objetivos propostos, como por exemplo a mudança do tipo da coluna "rental_date" na tabela "rental" para data com a localidade do Brasil, pois no formato data/hora que estava não era possível o relacionamento com a tabela "Data_tabela" criada;
- Criação de fórmulas e colunas com DAX e *M language*, com expressões como CALCULATE(), SUM(), AVERAGE(), COUNTROWS(), FILTER(), COALESCE(), IF(), ISBLANK(), BLANK(), CONCATENATE(), DIVIDE(), VAR, além da aplicação de indicadores e filtros;
- Criação de telas de background no Figma para melhor estética e usabilidade do usuário no Power BI;

Apresentar as análises que foram necessárias para o entendimento do negócio e que podem ser úteis no suporte para a tomada de decisão.

Em 2005 as vendas de DVD alcançaram R\$60,8 mil, destacando as categorias de filme “sports” e “Animation” com vendas no ano de R\$4892,19 e R\$4245,31, respectivamente. Entre os meses do ano com maior demanda, e consequente, vendas, salienta-se julho com 46% total anual. Em relação as locações, houve 15.862, com destaque com destaque para os filmes “Bucket Brotherhood” e para a atriz “Susan Davis”, com 34 e 816 locações, respectivamente. No mesmo ano o ticket médio foi de R\$4,22, destacando os filmes com participação de “Kirsten Paltrow” que alcançaram o maior ticket médio entre todos os atores, de R\$5,24. No que se refere aos clientes que ficaram com os DVDs por 9 dias, essa categoria representou a maior parte das vendas em comparação direta aos outros períodos alugados, com cerca de R\$10,4 mil em vendas. Entre os países, Austrália e Canadá apresentam a totalidade das vendas, com aproximadamente R\$30 mil cada. A taxa de perda do inventário - ou em outras palavras, os DVDs sem data de retorno -, ficou abaixo de 1%, o que representa somente 1 DVD não devolvido, representando excelente resultado.

No que se refere a 2006, que contempla vendas de DVD para somente fevereiro, o valor total alcançado foi de R\$514,18, o que representa ticket médio de R\$2,83 e 182 locações. Entre as categorias de filme com a maiores vendas destaca-se “Animation” e “Action” com R\$66,25 e R\$55,80, respectivamente. Em relação as locações, os filmes com o ator “Matthew Carrey” representam a maior participação com 14 alugueis, seguido por “Ewan Gooding” com 12 locações. Em relação as devoluções, dos 182 DVDs alugados, todos não apresentam data de retorno nos dados, o que coloca a taxa de perda de inventário em 100% para 2006. No mesmo ano na Austrália o ticket médio foi de R\$3,01 e houve 90 locações, já no Canadá ocorreu as 92 locações e ticket médio de R\$2,64.

Em relação aos clientes que apresentaram a maiores participações nas vendas da Discorama em 2005, destaca-se Eleanor Hunt (R\$211,55), Karl Seal (R\$208,58), Clara Shaw (R\$189,60) e Marcia Dean (R\$165,62). No que se refere a Eleanor Hunt que teve ticket médio de R\$4,70 e 46 locações no ano, cerca de 15% e 13% dos alugueis são das categorias “Sci-fi” e “Family”. No que se refere a categoria de filme “sports”, que representou o maior volume de vendas em 2005, destaca-se os clientes “Alex Gresham” e “Daryl Larue”, o primeiro com ticket médio de R\$4,70 e 7 locações, alcançando valor gasto nos alugueis de R\$32,93, já o segundo com 6 locações, ticket médio de R\$3,49 e valor gasto de R\$20,94.

Com base nas suas análises, uma conclusão com recomendações/sugestões de negócio e encaminhamentos para o CEO da Discorama.

- Adquirir mais filmes que representam a categoria “sports” e “Animation”, dado que demonstram grande volume de vendas e locações em 2005;
- Disponibilizar mais filmes com a atriz “Susan Davis”, a qual apresentou o maior volume de locações, e com o ator “Kirsten Paltrow”, que alcançou o maior ticket médio entre todos os atores em 2005;

- Proporcionar descontos e promoções para os clientes com as maiores participações nas vendas, e a criação de níveis com benefícios para os melhores clientes, como forma de manter a base de clientes e estimular o aluguel de novos filmes;
- Aplicar políticas de devoluções mais flexíveis e acordos de redução do valor por atraso, como forma de reduzir o indicador de taxa de perda do inventário em fevereiro de 2006.

Apresentar e justificar as ferramentas adotadas para a elaboração das análises.

Excel: Análise preliminar e entendimento dos dados;

Power Bi: criação de dashboards e indicadores dinâmicos;

Figma: Background com foco na usabilidade e estética do dashboard;

Python: Denormalizar duas tabelas para verificar e analisar a relação entre colunas, além de facilitar a análise por meio da construção de gráficos.

Github e git bash: criação do repositório e conexão para clonar o repositório.