

Gestor de BI

A PoD Bank acabou de criar o departamento de Business Intelligence e o gestor responsável pela área acabou de chegar na empresa.

A PoD Bank concede crédito para clientes com pouco histórico de crédito, portanto consome informações de bureaus e dados alternativos. Uma vez que a PoD Bank ainda não tem uma área de modelagem de crédito definida, a área de BI será responsável por atender o negócio e auxiliar na geração de insights e cenários que trarão respostas às estratégias de crescimento da empresa.

Para gerir as demandas de negócio que estão chegando o gestor precisa dos seguintes dados mapeados:

1. **Informações pessoais:** Nome, sexo, idade, estado civil, número de dependentes, ocupação, etc.
2. **Dados financeiros:** Renda do cliente, tipo de renda, valor do empréstimo solicitado, valor do empréstimo concedido, etc.
3. **Histórico de crédito:** Informações sobre pagamentos anteriores, atrasos de pagamento, montante de crédito utilizado, limites de crédito, etc.
4. **Informações de emprego:** Tipo de contrato de trabalho, tempo de emprego atual, categoria do trabalho, etc.
5. **Informações sobre imóveis:** Se o cliente possui ou não propriedade, tipo de propriedade, etc.
6. **Informações sobre educação:** Nível de educação do cliente, etc.
7. **Informações sobre o ambiente social:** Características do bairro onde o cliente reside, como índices de criminalidade, desenvolvimento humano, etc.
8. **Informações sobre redes sociais.**
9. **Informações sobre uso da internet:** sites navegados, compras na internet, o que houver de informações sobre navegação na internet.
10. **Informações sobre veículos.**
11. **Informações alternativas:** consumo de produtos de beleza, itens para pets...
12. **Informações sobre inadimplência, restritivos.**

Demanda 1) Mapeamento e catálogo dos dados que a empresa tem e se há necessidade de comprar informações externas.

Demanda 2) Disponibilizar os dados por assunto, de forma que seja fácil consumir indicadores de negócio, separados por assunto.

Sugestão: disponibilizar uma arquitetura com os dados históricos para uso imediato em BI e futuro em analytics

Demanda 3) Disponibilizar um catálogo de indicadores de negócio, que contenha:

- Taxa de Inadimplência: Calcula a porcentagem de clientes que estão inadimplentes em relação ao total de clientes.
- Valor Médio de Empréstimo: Calcula o valor médio dos empréstimos concedidos aos clientes.
- Taxa de Aprovação de Crédito: Mede a porcentagem de solicitações de crédito aprovadas em relação ao total de solicitações recebidas.
- Distribuição de Idade dos Clientes: Mostra a distribuição da idade dos clientes para identificar se há alguma faixa etária com maior risco de inadimplência.
- Taxa de Refinanciamento: Mede a porcentagem de clientes que refinanciam seus empréstimos em relação ao total de empréstimos concedidos.
- Tempo Médio de Atraso no Pagamento: Calcula a média de dias de atraso no pagamento dos clientes inadimplentes.
- Taxa de Recuperação: Avalia a porcentagem de empréstimos inadimplentes que foram recuperados através de ações de cobrança.
- Variação da Taxa de Inadimplência ao Longo do Tempo: Monitora as flutuações da taxa de inadimplência ao longo do tempo para identificar tendências e padrões.
- Performance de Segmentos de Clientes: Analisa a taxa de inadimplência e o valor médio de empréstimo por segmento de clientes (por exemplo, faixa de renda, localização geográfica, profissão/ocupação, escolaridade) para identificar grupos de maior risco e tomar ações direcionadas.
- Número de Créditos Ativos: Contabilize o número de créditos ativos de cada cliente, fornecendo insights sobre a carga de dívida e a capacidade de pagamento.
- Tipos de crédito que os clientes tomam no mercado, em volume e percentual

Esses indicadores devem estar disponíveis em um “data lake” prontos para consumo (façam sugestão de uma forma de consumo pra mim por favor).

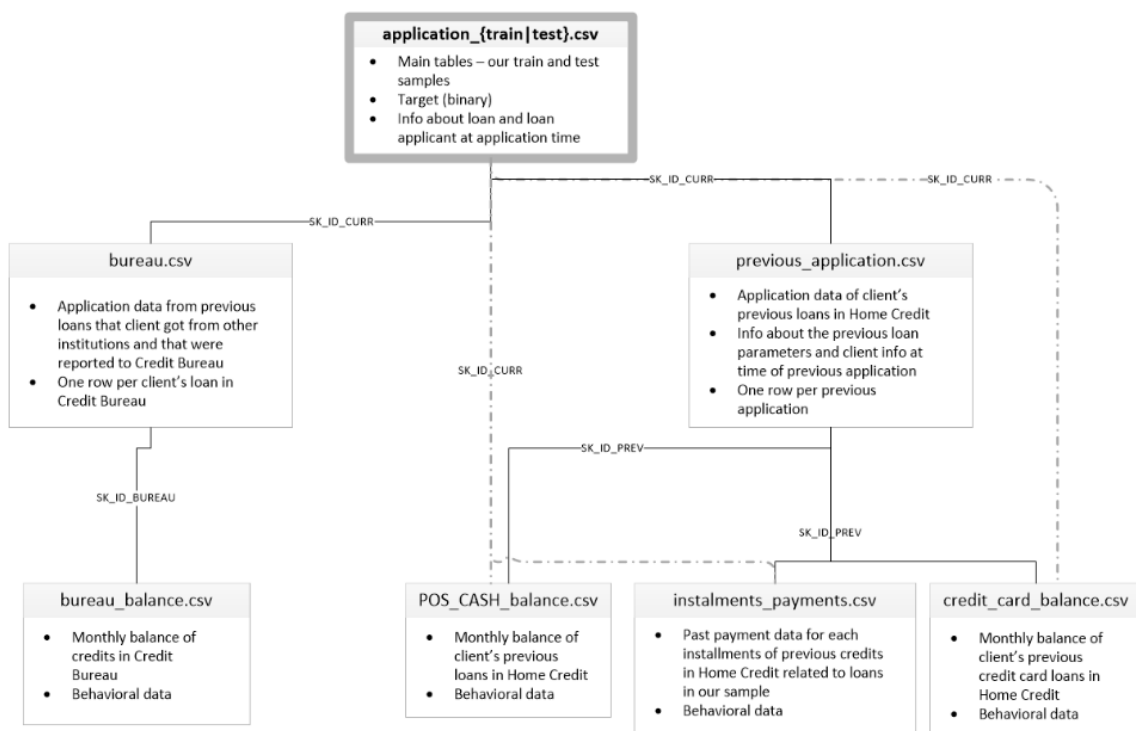
Demanda 4) Construção de um ou mais dashboards que contemple esses indicadores de negócio

Informações importantes

Os dados da empresa estão armazenados neste diretório:

https://s3.console.aws.amazon.com/s3/buckets/bkt-pod-academy-hackathon-data-eng?region=us-east-1&prefix=Diamantes_2023/PoDBank/Departamentos/Credito/&showversions=false

e temos este mapa de relacionamento de dados:



Sabendo que a tabela [HomeCredit_columns_description.csv](#) tem a descrição das tabelas e das variáveis presentes em cada tabela.

Sobre ferramentas que podem ser utilizadas:

Todas, mas o gestor quer um estudo dos custos que a solução vai trazer de acordo com a escolha das ferramentas que serão utilizadas. Este deve ser um dos entregáveis, um estudo apresentando todos os custos que a solução gera.

Dúvidas dos Alunos PoD Academy



Gabriela Andrechuk Feniman 14/07/2023 21:57

Boa noite, pessoal!

Fizemos uma análise das bases disponibilizadas e temos 4 dúvidas de largada:

1. Na tabela "application_test" disponibilizada, o que seriam os documentos mencionados nos campos FLAG_DOCUMENT_XX?
[@Juliana.zucoloto](#) / [@Dalmer Sella](#)
2. Na tabela "application_test" disponibilizada, o que seriam as informações mencionadas nos campos EXT_SOURCE_X? [@Juliana.zucoloto](#) / [@Dalmer Sella](#)
3. Na tabela "application_test" disponibilizada, está faltando o campo TARGET mencionado na tabela "HomeCredit_columns_description". Esse campo deveria estar na primeira tabela também? [@Juliana.zucoloto](#) / [@Dalmer Sella](#)
4. Sabemos quais dados precisaríamos comprar. A gente deve fazer uma solicitação formal pedindo esses dados? Ou é algo que devemos só mencionar na hora de apresentar o case?
[@Juliana.zucoloto](#) / [@tatioli](#)



ednamanhu Ontem às 22:13

@Admin - PoD Academy, boa noite.

A resposta sobre a aprovação do crédito solicitado tem qual SLA. Ou seja, até 2 dias, 3 dias, 24 horas?



ednamanhu Hoje às 00:21

@Admin - PoD Academy na tabela

Installments_payments, como devemos entender a Variável "Version of installment calendar (0 is for credit card) of previous credit. Change of installment version from month to month signifies that some parameter of payment calendar has changed". Por que há a menção de cartão de crédito? Não se trata das parcelas do financiamento?

Variável: "When the installment of previous credit was supposed to be paid (relative to application date of current loan)" a contagem de dias é em relação à data de vencimento escolhida no momento da abertura do crédito?

Variável: "What was the prescribed installment amount of previous credit on this installment" está como Amount. Devemos entender que é o cálculo da parcela atrasada com os devidos encargos?



ednamanhu Ontem às 23:39

@Admin - PoD Academy na tabela Bureau, como devemos entender a variável "How many times was the Credit Bureau credit prolonged""



ednamanhu Ontem às 23:25

@Admin - PoD Academy na tabela Application_test, como devemos entender a variável "How many days before the application did client change his registration""

@Admin - PoD Academy Como devemos entender o conceito de "receivable" da tabela credit_card_balance, conforme o descritivo das variáveis:

- Amount receivable for principal on the previous credit;
- Amount receivable on the previous credit;
- Total amount receivable on the previous credit;



ednamanhu Ontem às 22:49

@Admin - PoD Academy ,

"How much did the client pay during the month in total on the previous credit" da tabela credit_card_Balance está como Amount_payment_total_current. Esta variável está correta? Realmente é valor ou deveria ser uma relação entre valor pago/valor total?



ednamanhu Ontem às 22:42

@Admin - PoD Academy por favor, qual a diferença entre estas variáveis:

Amount drawing during the month of the previous credit

Amount of other drawings during the month of the previous credit

ambas fazem parte da tabela
credit_card_Balance

Informações de Negócio sobre os dados disponibilizados:

application_{train|test}.csv:

Esta tabela contém pessoas que já tomaram crédito e honraram ou quitaram ou não o contrato.

Tabela principal, também conhecida como Tabela Analítica de Modelagem, dividida em dois arquivos para Treinamento (com a variável TARGET) e Teste (sem a variável TARGET).

Contém dados estáticos de todas as solicitações de empréstimo. Cada linha representa um empréstimo em nossa amostra de dados.

bureau.csv:

Contém informações sobre todos os créditos anteriores do cliente fornecidos por outras instituições financeiras que foram relatados ao Bureau de Crédito (para clientes que possuem empréstimo em nossa amostra).

Para cada empréstimo em nossa amostra, existem várias linhas que correspondem ao número de créditos que o cliente teve no Bureau de Crédito antes da data da solicitação.

bureau_balance.csv:

Apresenta os saldos mensais dos créditos anteriores relatados ao Bureau de Crédito.

Essa tabela tem uma linha para cada mês da história de cada crédito anterior relatado ao Bureau de Crédito - ou seja, a tabela tem (#empréstimos na amostra * #de créditos anteriores relativos * #de meses em que temos algum histórico observável para os créditos anteriores) linhas.

POS_CASH_balance.csv:

Contém os saldos mensais dos créditos anteriores relacionados a vendas a varejo (POS) e empréstimos em dinheiro que o solicitante teve com a Home Credit.

Essa tabela tem uma linha para cada mês da história de cada crédito anterior relacionado à Home Credit (créditos ao consumidor e empréstimos em dinheiro) em relação aos empréstimos em nossa amostra - ou seja, a tabela tem (#empréstimos na amostra * #de créditos anteriores relativos * #de meses em que temos algum histórico observável para os créditos anteriores) linhas.

credit_card_balance.csv:

Apresenta os saldos mensais de cartões de crédito anteriores que o solicitante tinha com a Home Credit.

Essa tabela tem uma linha para cada mês da história de cada cartão de crédito anterior relacionado à Home Credit em relação aos empréstimos em nossa amostra - ou seja, a tabela tem (#empréstimos na amostra * #de cartões de crédito anteriores relativos * #de meses em que temos algum histórico observável para os cartões de crédito anteriores) linhas.

previous_application.csv:

Contém informações sobre todas as solicitações de empréstimo anteriores dos clientes que têm empréstimos em nossa amostra.

Há uma linha para cada solicitação de empréstimo anterior relacionada aos empréstimos em nossa amostra de dados.

installments_payments.csv:

Apresenta o histórico de pagamentos para os créditos anteriormente concedidos pela Home Credit relacionados aos empréstimos em nossa amostra.

Existem a) uma linha para cada pagamento efetuado e b) uma linha para cada pagamento perdido.

Uma linha é equivalente a um pagamento de uma parcela OU uma parcela correspondente a um pagamento de um crédito Home Credit anterior relacionado aos empréstimos em nossa amostra.

HomeCredit_columns_description.csv:

Este arquivo contém a descrição das colunas presentes nas tabelas acima, fornecendo informações detalhadas sobre os dados em cada tabela.