

PUC-RIO

MVP SPRINT ENGENHARIA DE DADOS

Análise do aplicativo Gelato

ANA LUIZA ABRAO RORIZ SOARES DE CARVALHO

## MOTIVAÇÃO

A motivação deste MVP vem da necessidade de analisar os dados de uso e popularidade de features do Gelato, um aplicativo para gestão de obras. O Gelato foi criado em 2024 pela autora deste trabalho juntamente com uma sócia arquiteta. Contratou-se uma “software house” que o desenvolveu em Bubble, uma ferramenta “no-code”. As criadoras não tiveram qualquer controle a modelagem de banco de dados usada ou quaisquer questões técnicas à época da criação.

O Gelato é um aplicativo para gestão de obras de construção ou reformas, usado por arquitetos, engenheiros e proprietários para acompanharem a transformação e os trabalhos realizados no imóvel. Ele é bem mais simples, visual e amigável que um ERP de engenharia e permite o acompanhamento da obra através do upload de fotos da obra, registro das pendências a serem resolvidas e um caderno de notas para registro de fatos ou decisões. Estes dados ficam salvos no aplicativo e podem ser acessados pela equipe de profissionais e pelo cliente final.

Na criação do Gelato optou-se por fazer uma versão inicial simples em uma ferramenta no-code e disponibiliza-lo com uma versão gratuita bastante generosa em funcionalidade e armazenamento. Com isso buscava-se entender quais funcionalidades seriam úteis no dia a dia do acompanhamento de obras. Espera-se que em sua próxima versão o Gelato seja lançado como um aplicativo nativo, com maior custo de desenvolvimento, portanto as criadoras tem a intenção de investir apenas nas funcionalidades que mostram tração com os usuários.

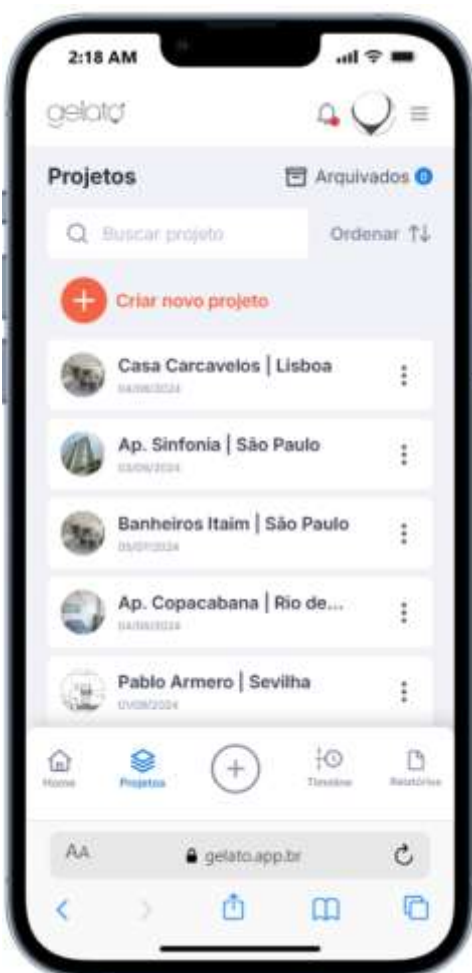
A motivação principal deste trabalho é fazer a primeira análise do uso do aplicativo, entender o engajamento dos usuários e seu comportamento ao longo do tempo e consolidar informações provenientes de diferentes módulos do aplicativo (projetos, visitas, fotos, pendências, ambientes e serviços). Aqui tem-se o objetivo estratégico de usar os dados coletados em apoiar a tomada de decisão em três frentes:

- a) Avaliação do valor entregue pelo aplicativo ao identificar padrões reais de uso e funcionalidades mais importantes
- b) Prioridade de evolução e desenvolvimento do produto ao fundamentar decisões sobre quais áreas devem receber investimento na próxima versão
- c) Monitoramento do comportamento do usuário ao permitir acompanhamento histórico, comparações, tendências e possíveis gargalos.

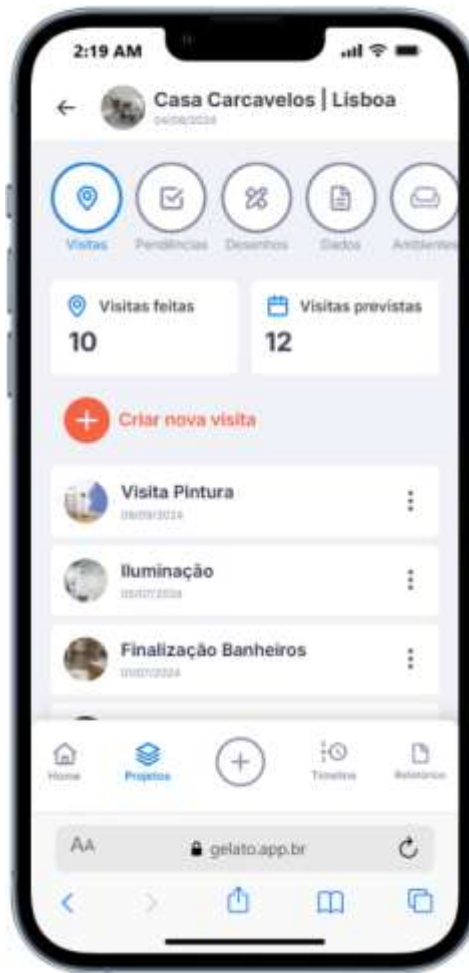
## O APLICATIVO

A fim de dar melhor clareza sobre os dados e os resultados obtidos através dele, é importante explicar brevemente o funcionamento do aplicativo.

A primeira coisa que o usuário deve criar é um projeto. Neste projeto o usuário pode dar mais detalhes sobre a obra como ambientes que sofrerão intervenção e serviços a serem executados. Também é possível subir arquivos como plantas, contratos e outro documentos referentes àquele projeto. Dentro do projeto o usuário deve criar visitas, que são onde os principais dados do projeto “moram”.

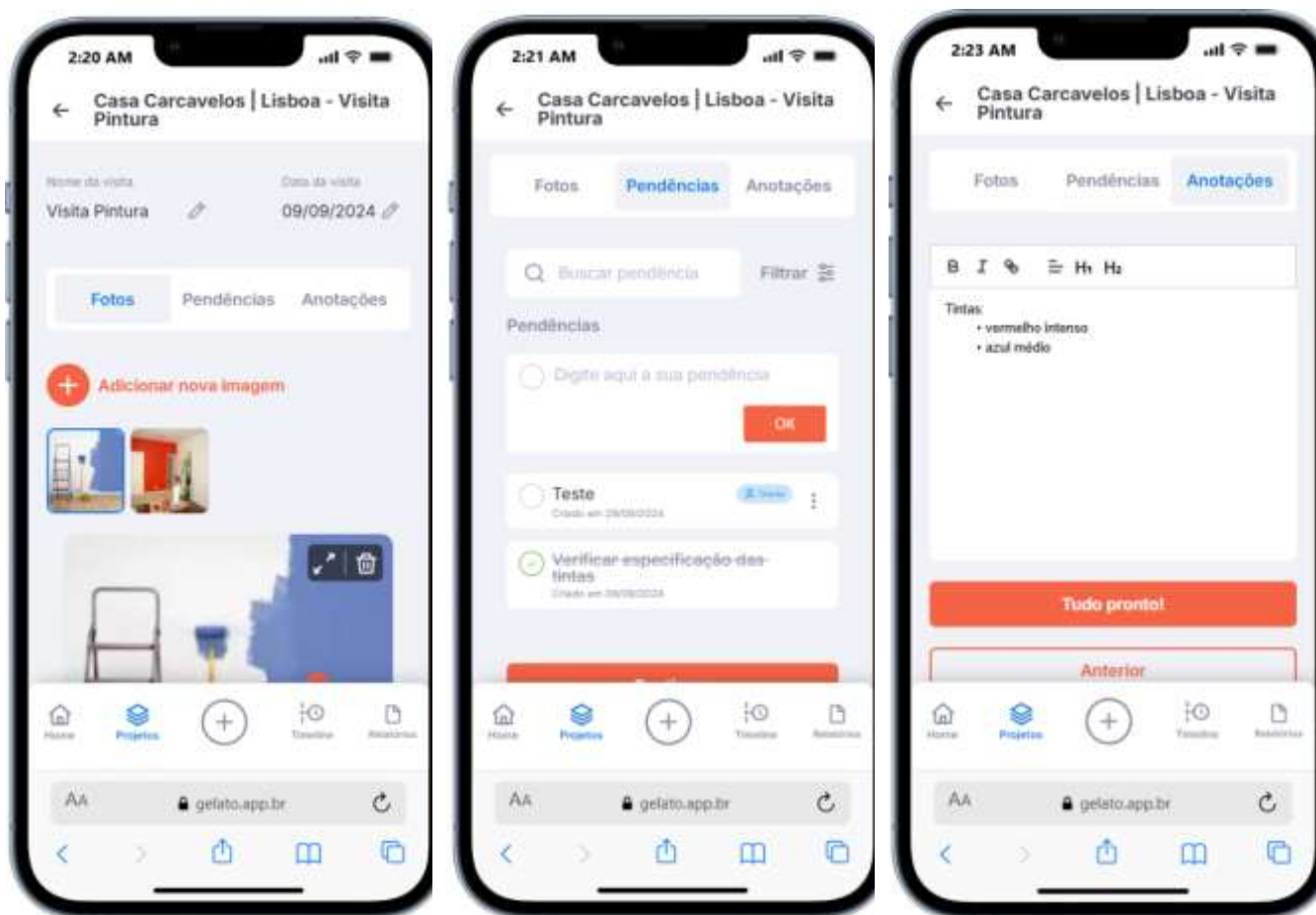


Tela da listagem de projetos de um usuário



Tela da listagem das visitas dentro de um projeto

O Gelato foi criado pensando em como o arquiteto acompanha obra, logo seu principal ponto de contato com o usuário é a visita à obra. Cada vez que o arquiteto vai à obra, ele cria uma nova visita dentro de um determinado projeto. Ali, dentro daquela visita, ele insere fotos, anotações e pendências, pode marcar as pendências como urgentes e pode marcar as fotos como visíveis ao cliente. Além disso o usuário também pode colocar etiquetas nas fotos. Estas etiquetas podem ter qualquer texto, mas como default o sistema oferecerá como etiquetas os ambientes e serviços adicionados ao projeto.



Os dados do aplicativo se encontram em tabelas dentro do banco de dados da plataforma Bubble e podem ser exportadas em arquivos csv. Essas tabelas incluem informações sobre usuários, projetos, visitas, fotos, pendências, ambientes e serviços, além de dados auxiliares como escritórios e equipes. Cada tabela armazena os registros conforme criados pelos usuários no aplicativo, utilizando identificadores próprios gerados pelo Bubble e campos predominantemente textuais.

## MODELAGEM

### Os requisitos do negócio

Seguindo o material do curso disponibilizado em (Baião & Lifschitz, 2025) o processo de desenvolvimento de um DW pressupõe as seguintes etapas: (a) definição do escopo; (b) levantamento de requisitos e análise do problema; (c) projeto do DW; (d) integração de dados e testes; (e) implantação.

Como a definição do escopo foi coberta na sessão de Motivação, vou me ater ao levantamento de requisitos e análise do problema e ao projeto do DW nesta sessão.

O levantamento de requisitos consiste em identificar quais informações são necessárias para apoiar as decisões de negócio, convertendo objetivos estratégicos em medidas quantificáveis — exatamente como recomendado por (Baião & Lifschitz, 2025) ao explicar que um ambiente analítico deve ser “orientado ao assunto” e permitir análise histórica dos fatos relevantes do negócio. Portanto, as medidas de interesse para monitoramento do Gelato serão indicadores de desempenho contabilizadas com base em fatos históricos de engajamento dos usuários com a plataforma.

Para chegar a estas medidas usaremos as seguintes perguntas de negócio com respeito à interação dos usuários do Gelato:

- 1) Qual é o número de usuários ativos e a frequência de uso?
- 2) Quantos projetos e visitas são criados por usuário?
- 3) Quantas fotos são adicionadas e como se distribuem por visita ou projeto?
- 4) Qual é o volume de pendências criadas e sua urgência?
- 5) Quais ambientes e serviços personalizados são mais utilizados?
- 6) Qual o uso da funcionalidade anotação dentro das visitas?

Para responder estas perguntas, são necessárias medidas numéricas que podem ser analisadas ao longo do tempo. Em modelagem dimensional essas medidas são representadas por  **fatos**  que correspondem a eventos feitos pelos usuários no Gelato. Com base nas perguntas de negócios, há três eventos centrais que representam ações fundamentais dos usuários na plataforma:

- a) Criação de visitas
- b) Upload de fotos
- c) Criação de pendências

Dados estes eventos, podemos definir três fatos, ou medidas de interesse que devem ser monitoradas, para o Gelato como eventos atômicos feitos pelos usuários na plataforma que posteriormente podem ser agregados:

**FATO 1) Cada visita criada em um projeto**

**FATO 2) Cada foto adicionada a uma visita**

**FATO 3) Cada pendência criada**

Estes três fatos têm forte ligações as dimensões do usuário, projeto, ambientes, serviços e tempo dentro do aplicativo. Essas características e a necessidade de responder pergunta de negócio que envolvem contagens, frequências e funis fazem com que o Modelo Estrela seja a ferramenta mais adequada para o modelo lógico do Data Warehouse.

Para o item “(c) Projeto do DW”, conforme descrito em (Baião & Lifschitz, 2025), serão implementadas as tabelas de fato e dimensão no SGDB relacional Databricks. Ou seja, será usada uma abordagem ROLAP (Relational Online Analytical Processing), onde cada fato e dimensão do esquema será implementado em uma tabela.

A descrição das tabelas e os atributos dos fatos podem ser encontrados no Anexo do Catálogo dos dados.

## Os dados

O Gelato é um sistema para uso no processo de acompanhamento de obras, em que o usuário insere no sistema as informações que ele julga serem relevantes, e posteriormente compartilha com pessoas interessadas na obra ou as usa para seu próprio controle da evolução dos trabalhos. Ele é orientado à execução rápida das transações e voltado para a perspectiva operacional da construção ou reforma, podendo ser classificado como um sistema de *on-line transactional processing* (OLTP).

Em sua estrutura atual os dados são guardados em tabelas no próprio sistema Bubble e divididos nas seguintes tabelas: **usuários, projetos, visitas, fotos, pendências, ambientes e serviços**. A forma como estas tabelas estão estruturadas não atendem às necessidades analíticas do projeto. É possível extrair informações de cada uma das tabelas, mas não é possível cruzar informações entre elas de forma fácil e automática.

Na Figure 1- Estatísticas da tabela Fotos do aplicativo Gelato para o dia 6/12/2024 Figure 1 mostrada abaixo é possível ver as informações da **tabela foto** mostrando diversas métricas retiradas automaticamente da base. Porém não é possível manipular os dados desta tabela de forma direta e automática para quebrar estas mesmas estatísticas por usuários ou projetos, nem mesmo fazer uma série histórica com estas estatísticas dia a dia sem extensas manipulações.

Fotos	Visitas com fotos	468
	% Visitas com fotos	76%
	Fotos por visita	8.3
	Fotos	3893
	Marcação na foto	86
	% Fotos com Marcação	2%
	Etiquetas nas fotos	2141
	% Fotos com etiqueta	55%
	Fotos pro Cliente	210
	% Fotos pro Cliente	5%

Figure 1- Estatísticas da tabela Fotos do aplicativo Gelato para o dia 6/12/2024

Outro exemplo pode ser visto na Figure 2. Os dados extraídos da **tabela pendências** mostram que as pendências podem ser agregadas de acordo com as colunas presentes na tabela, mas não é possível de imediato analisar de quais projetos ou visitas elas vêm, muito menos ver um padrão entre o uso ou não de pendências entre os usuários.

Pendências	Total	408
	Central	177
	Visita	231
	% Central	43%
	% Visita	57%
	% Urgentes	20%
	% Clientes	13%
	% Resolvidas	21%

Figure 2- Estatísticas da tabela Pendências do aplicativo Gelato para o dia 6/12/2024

Os exemplos reais das dificuldades no dia a dia do Gelato mostram de forma muito clara os obstáculos que a ausência de uma estrutura multidimensional em um sistema OLTP trazem para a tomada de decisões por parte da gerência de um negócio. No Bubble, assim como em muitos sistemas OLTP os dados ficam fragmentados em múltiplas tabelas e com chaves fracas, impedindo análises agregadas, cruzamentos consistentes e principalmente análises históricas.

Existe uma necessidade em transformar os dados extraídos do sistema para poder conduzir análises mais completas que levem a conclusões que possam ajudar a tomada de decisão no negócio, reforçando a necessidade de um DW estruturado em fatos e dimensões para conduzir as análises propostas.

Depois que os **fatos** foram especificados na sessão anterior de requisitos do negócio, deve-se prosseguir para determinar as **dimensões**, ou características, que são necessárias para contextualizar cada um deles e também os atributos do fato em si.

As dimensões que usadas para caracterizar nossos três fatos serão as seguintes:

**DIMENSÃO 1) Usuário**

**DIMENSÃO 2) Projeto**

**DIMENSÃO 3) Visita**

**DIMENSÃO 4) Tempo**

**DIMENSÃO 5) Pendência**

A escolha destas dimensões reflete o funil de elementos presentes no Gelato até a criação dos fatos. A dimensão do Usuário caracteriza quem executa as ações e o primeiro passo de contato de alguém com o aplicativo. A dimensão Projeto caracteriza a primeira ação que o usuário pode fazer dentro do sistema, que é a criação de um projeto dentro do qual as visitas irão acontecer. Logo a dimensão Visita descreve a entidade onde fotos, anotações e pendências se relacionam. O Tempo permite análises temporais e finalmente a dimensão Pendência normaliza as descrições do texto das pendências criadas, já que os atributos operacionais da pendência já estão armazenados em seu fato correspondente.

A descrição completa das dimensões, suas tabelas e atributos podem ser encontrados no Anexo Catálogo dos dados. O formato final do Modelo Estrela do Data Warehouse para o Gelato é dado pela abaixo.





Figura 1 - Modelo Estrela do DW do Gelato

## EXTRAÇÃO, CARGA E TRANSFORMAÇÃO DOS DADOS

Após o modelo ter sido definido, partiu-se para a modelagem dos dados através da plataforma Databricks.

O primeiro passo foi fazer o *upload* dos arquivos originais (csv) na plataforma e depois disso seguimos para a criação das tabelas bronze, prata e ouro, como pode ser visto nos notebooks MVP\_etl\_bronze, MVP\_etl\_silver, MVP\_etl\_gold.

Nas tabelas bronze temos o início do processo de extração, transformação e carga dos dados (ETL na sigla em inglês). Bronze: criou-se tabelas bronze\_users, visitas, etc. quase tudo estava no formato string após a importação dos dados.

O passo seguinte foi criar as tabelas prata. Esta etapa existe para transformar os dados crus e inconsistentes (como por exemplo as datas que estão como string) em dados limpos, padronizados e prontos para modelagem. Nesta etapa vamos corrigir problemas nos dados,

padronizar seus tipos, tratar duplicidades e preparar os dados para a camada ouro, que será o modelo estrela explicado anteriormente. Os comandos utilizados podem ser consultados no arquivo MVP\_etl\_silver disponível no GitHub.

Após o processo de ETL ter sido concluído, foi executada a fase das tabelas ouro, que será a implementação do modelo estrela descrito anteriormente. Os dados nela são organizados de acordo com os fatos e suas dimensões especificadas para o Gelato. Criou-se primeiramente as dimensões do modelo, a partir das tabelas prata e posteriormente, foram feitas várias operações de JOIN para a criação das tabelas fato, como pode ser visto nos códigos do notebook disponível na plataforma GitHub MVP\_etl\_gold.

A maior dificuldade do processo foi criar as tabelas fato, pois temos muitas dimensões no modelo, como foi explicado na sessão anterior. O comando JOIN e o LEFT JOIN foram os principais comandos usados nesta fase do trabalho.

## AS ANÁLISES

### Qualidade dos dados

Antes de iniciar as análises propriamente ditas deve-se comentar sobre a qualidade dos dados. A análise da qualidade dos dados foi realizada ao longo de todo o processo de ETL, abrangendo as camadas *bronze*, *silver* e *gold*. Para cada atributo relevante, foram avaliados aspectos como presença de valores nulos, inconsistências de formato, padronização de tipos de dados e duplicidade de registros.

Na camada bronze, os dados foram ingeridos diretamente a partir de arquivos CSV exportados da plataforma Bubble, sem qualquer tratamento prévio, preservando fielmente a estrutura original. Nessa etapa, identificaram-se diversos problemas típicos de dados operacionais, tais como nomes de colunas com caracteres especiais, espaços, acentuação e inconsistência de nomenclatura (por exemplo, “unique id”, “TagServiço”, “Observação”). Esses problemas inviabilizam o uso direto das tabelas em um ambiente analítico e foram tratados posteriormente. Também foi observada a presença de campos textuais com valores vazios ou nulos, além de datas armazenadas como *string*, muitas vezes contendo informações de data e hora em formatos heterogêneos.

Na camada silver, realizou-se o processo de limpeza, padronização e conformação dos dados. Os atributos foram renomeados seguindo um padrão *snake\_case*, as datas foram

convertidas para o tipo DATE. O formato das datas foi um problema nesta fase. Inicialmente esperava-se poder analisar data e hora de diversas ações, mas o formato em que o Bubble gravava os dados de data e hora geraram muita dificuldade no processo de ETL. Logo escolheu-se deixar o timestamp e trabalhar com apenas a data nas tabelas.

Durante a análise exploratória, constatou-se que alguns atributos, embora presentes no esquema, não continham valores preenchidos em nenhum registro. Esse foi o caso dos campos relacionados à categorização por ambiente e serviço (`tag_ambiente`, `tag_servico` e `ambiente_servico_tag`). A verificação foi feita por meio de contagens de valores não nulos e consultas de valores distintos, que retornaram zero ocorrências válidas. Isso inviabilizou a resposta da pergunta 5.

Nas tabelas gold os dados foram organizados em um modelo dimensional, composto por tabelas fato e tabelas de dimensão, utilizando chaves substitutas (*surrogate keys*). Nessa etapa, a qualidade dos dados foi verificada por meio da validação dos relacionamentos entre fatos e dimensões, garantindo que não existissem inconsistências entre as tabelas. Situações como projetos sem visitas ou usuários sem registros de atividade foram tratadas explicitamente nas análises, utilizando *LEFT JOINS*

## Solução do problema

Com as tabelas do modelo estrela prontas, partiu-se para responder as perguntas de negócio colocadas na sessão “Os requisitos do negócio”.

**É importante notar que muitas análises dizem respeito a quais usuários fizeram determinada ação, mas para manter a privacidade dos usuários (que no caso do Gelato são identificados pelo seu e-mail) optou-se por não incluir as tabelas finais neste trabalho. Porém os códigos podem ser vistos nos notebooks divulgados.**

### 1) Número de usuários ativos e frequência de uso

Definiu-se como usuários ativos aqueles que criaram pelo menos uma pendência ou foto dentro de uma visita. Para os códigos referentes a essa pergunta utilizou-se o comando COUNT, como pode ser visto nos códigos no notebook MVP\_analise\_pergunta\_1.

Há 76 usuários ativos na base, de um total de 706. Ou seja apenas 10.76% dos usuários conseguiu criar tarefas básicas no Gelato, mostrando um problema de onboarding ou UX/UI.

Além de uma lista com os usuários mais ativos, que é valiosa para poder abordá-los e entender como o aplicativo poderia melhorar, houve um insight valioso sobre a frequência de uso. Internamente no Gelato acreditávamos que a segunda feira deveria ser o dia com mais

engajamento, pois seria o dia em que os arquitetos procurariam ir a obra e resolver suas pendências no início da semana. Inclusive focávamos ações de marketing nesse dia, mas para nossa surpresa o dia com mais uso é a sexta-feira como pode ser visto na tabela abaixo. A segunda-feira se mostrou um dia com baixa relevância de uso.

	$A_0^B$ dia_semana	$A_3^B$ visitas
1	Friday	200
2	Wednesday	158
3	Tuesday	125
4	Thursday	105
5	Monday	63
6	Sunday	16
7	Saturday	9

Para essa análise foi imprescindível ter criado a dimensão tempo no modelo estrela.

Também analisamos os dias em que os usuários mais adicionam fotos e a sexta-feira seguiu o padrão de ter mais fluxo de usuários.

Finalmente também pode-se ver os usuários mais ativos no aplicativo, mas para guardar a confidencialidade dos dados optou-se por não divulgar esta tabela aqui.

## 2) Quantos projetos e visitas são criados por usuário

Para a segunda pergunta buscou-se descobrir o perfil de criação de projeto dos usuários, já que este é o primeiro passo no funil para uso do aplicativo.

Foi feito o seguinte histograma que mostra que a grande maioria dos usuários tem interesse no aplicativo mas não consegue ou não quer criar nenhum projeto. Isso nos dá o insight de que o processo de criação de projetos pode estar muito longo ou complicado.

	$A_0^B$ faixa	$A_3^B$ usuarios_nessa_faixa
1	1 projeto	49
2	10+ projetos	18
3	2 a 4 projetos	24
4	5 a 10 projetos	7
5	Nenhum projeto	608

Há 6,9% dos usuários com apenas 1 projeto e 3.3% dos usuários com 2, 3 ou 4 projetos. Isso nos mostra que poucos usuários completam a criação de um projeto, mas caso o façam há algum tipo de engajamento posterior. Há um número relativamente alto de usuários com mais de 10 projetos mas isso é explicado pela promoção concierge, em que durante um período ofereceu-se uma

assistente que preencheria o sistema pelo usuário. Os usuários desta promoção usaram muito o Gelato, o que mostra que há interesse em guardar os projetos mas há uma barreira em preencher o sistema, o que nos reforça que a experiência de UX/UI está aquém do potencial.

Para ver o padrão da criação de visitas criou-se um histograma parecido e percebeu que há uma barreira grande a criar a primeira visita. Há 26.7% dos usuários que não criam nenhuma visita. Eles chegam a criar projetos mas não percebem ou sabem que o passo seguinte é a criação da visita. Quando ultrapassada a barreira da criação da primeira visita, temos que 54% dos usuários que criaram a primeira visita continuam para criar a segunda ou a terceira. Isso nos mostra que há dificuldade com o uso inicial do aplicativo, mas quando essa barreira é vencida o uso melhora.

	A <sup>B</sup> <sub>C</sub> faixa	1 <sup>2</sup> <sub>3</sub> projetos_nessa_faixa
1	a) 0 visitas	189
2	b) 1 visita	135
3	c) 2 a 3 visitas	73
4	d) 4 a 5 visitas	25
5	e) 6 a 10 visitas	17
6	f) 11 a 20 visitas	7
7	g) 20+ visitas	1

Depois dessa contagem partiu-se para entender quais são os usuários que mais criam visitas e o número de visitas por projetos destes usuários, mas por motivos de privacidade escolheu-se omitir essa tabela. O que pode ser mostrado abaixo é o histograma com a distribuição dos usuários por número de visitas criadas. Também pode-se ver que o número médio de visitas por projeto é de 2.5 e o máximo de visitas em um projeto é de 32.

	A <sup>B</sup> <sub>C</sub> faixa	1 <sup>2</sup> <sub>3</sub> usuarios_nessa_faixa
1	0 a 5 visitas	75
2	21 a 50 visitas	6
3	50+ visitas	3
4	6 a 20 visitas	14

	1 <sup>2</sup> <sub>3</sub> min_visitas	1.2 avg_visitas	1 <sup>2</sup> <sub>3</sub> max_visitas
1	1	2.54263565891472...	32

### 3) Quantas fotos são adicionadas e como se distribuem por visita ou projeto?

Para a pergunta 3 explorou-se o padrão de upload de foto no aplicativo. Inicialmente foi feita uma contagem e viu-se que há 37.730 fotos na base de dados do Gelato.

Estas fotos estão presentes em 75% das visitas. Isso mostra que a funcionalidade de upload de fotos é amplamente usada pelos usuários que conseguem quebrar as barreiras de criar projeto e visita. Ou seja, o maior valor que o Gelato entrega está na gestão das fotos da obra.

	.00 percentual_visitas_com_foto
1	75.28

Analisando as visitas que mais tem fotos adicionadas, percebeu-se que há um padrão nos nomes. Visitas de antes e depois são as mais documentadas no aplicativo. Há muitas visitas com um padrão de nome que remete ao fim de uma obra e também à fase anterior à obra como vistas de antes, levantamento ou 3D.

	A <sup>B</sup> <sub>C</sub> nome_visita	1 <sup>2</sup> <sub>3</sub> qtd_fotos
1	Finalizado	15540
2	3D	11430
3	Obra Pronta	3728
4	Antes	1040
5	Finalizado	960
6	Antes	518
7	3D	306
8	Obra concluída	294
9	Obra – Demolição	208
10	Demolição	155
11	Obra pronta	153
12	Obra - Pintura	115
13	23 agosto 2023	114
14	Levantamento	110
15	Vistoria Construtora	82

Partiu-se depois para ver quais projetos mais tinham fotos adicionadas (omitido aqui por privacidade) e os usuários que mais subiram fotos (também omitido por privacidade).

Avaliou-se a feature de marcação das fotos, em que o usuário tem a opção de colar uma anotação na foto. Essa feature depende de um plugin externo no aplicativo e deixa o processo de upload mais lento. Percebeu-se que a esmagadora maioria das fotos não usa essa feature, o que nos dá

confiança em abandoná-la para agilizar o processo de upload. Também pudemos identificar o usuário que mais usou essa feature a fim de avisá-lo sobre um possível fim desta.

	☰ foto_tem_marcaador	1 <sup>2</sup> 3 total
1	false	37488
2	true	242

Finalizou-se essa sessão analisando algumas estatísticas sobre a distribuição das fotos nos projetos e visitas. Pode-se ver que há um padrão similar ao da criação das visitas. Há muitos projetos com nenhuma ou poucas fotos, mostrando a dificuldade do usuário em entender como usar a feature, mas há um engajamento satisfatório para um uso razoável de fotos. Há 19,3% das visitas com 6 a 11 fotos que é um uso razoável. Também pode-se ver que há um certo abuso no número de visitas com mais de 50 fotos, muito provavelmente vindo dos usuários que tiveram acesso à assistente para preencher o sistema por eles. Os usuários tinham apenas que enviar as fotos por WhatsApp para que a assistente as colocasse no sistema. Entende-se que estes usuários puxaram a média de fotos por visita e pro projeto.

	A <sup>B</sup> C faixa	1 <sup>2</sup> 3 quantidade_de_visitas
1	0 a 5 fotos	203
2	6 a 11 fotos	110
3	12 a 20 fotos	35
4	21 a 35 fotos	43
5	36 a 50 fotos	13
6	50+ fotos	138

	.00 media_fotos_por_projeto
1	189.60

	.00 media_fotos_por_visita
1	69.61

#### 4) Qual é o volume de pendências criadas e sua urgência?

Para responder essa pergunta começou-se por contar o total de pendências criadas e a quantidade daquelas que são urgentes.

	$1^2_3$ total_pendencias
1	542

	$\Sigma$ eh_urgente	$1^2_3$ total
1	true	156
2	false	386

Pode-se ver que 28,78% das pendências criadas são urgentes. Além disso pode-se ver que o percentual de visitas com pendências criadas ficou abaixo de 25%, o que foi uma decepção. Esperava-se que mais usuários utilizassem essa funcionalidade.

	.00 percentual_visitas_com_pendencias
1	24.58

Posteriormente foi feito o levantamento do número de pendências por projeto e por usuário, que não são reportadas aqui por motivo de proteção de dados, mas percebeu-se uma forte concentração no uso da feature pendências. Apenas 5 usuários criaram 46,86% das pendências no aplicativo e que se adicionarmos os 5 principais usuários seguintes, vemos que apenas 10 usuários criaram 66% das pendências, mostrando que a maioria dos usuários ativos não explorou a ferramenta.

Nota-se que 65% das pendências cadastradas não foram resolvidas até o corte da base de dados. Mesmo os usuários engajados que criaram pendências não voltam para marca-las como solucionadas, considerando que as pendencias devem ser resolvidas para a entrega de uma obra.

	$\Sigma$ resolvida	$1^2_3$ total
1	true	186
2	false	356

Encerramos a análise das pendências com uma constatação curiosa. Apesar de sexta-feira ser o dia de maior número de upload de fotos e criação de projetos, a terça-feira é o principal dia para a criação das pendencias.

	$\Delta^B_C$ dia_semana	$1^2_3$ total
1	Tuesday	154
2	Wednesday	110
3	Thursday	87
4	Friday	65
5	Saturday	55
6	Monday	52
7	Sunday	19



Isso de certa forma mostra que os usuários podem ir na obra no início da semana) porém na terça e não na segunda-feira como esperava-se), mas eles usam o aplicativo na obra apenas para anotações de texto. O upload das fotos é feito no fim da semana como se fosse uma rotina de encerramento da semana. Outro motivo para isso acontecer é o fato do carregamento de fotos não ter a velocidade ideal por causa do plugin da marcação nas fotos como mencionado anteriormente.

##### 5) Quais ambientes e serviços personalizados são mais utilizados?

Esta pergunta não pode ser respondida porque as bases exportadas do Bubble vieram vazias, como foi comentado na sessão anterior de qualidade dos dados. Houve alguma decisão de mudar a base de dados por parte da software house que criou o sistema de salvar os serviços e ambientes personalizados em tabelas diferentes das que estavam marcadas pelo título “ambiente” e “serviço”. A exploração desta decisão e da estruturação destes campos no aplicativo requereriam um tempo extra não condizente com os prazos deste trabalho.

##### 6) Qual o uso da funcionalidade anotação dentro das visitas?

A última pergunta trata da funcionalidade anotação dentro da visita. Esta funcionalidade foi criada com o intuito de servir como uma ata do que acontecia na visita. Esperava-se que o usuário fosse escrever itens que não eram pendências e sim constatações ou decisões tomadas durante a visita à obra, como por exemplo “Escolheu-se a cor XYZ da Suvinil para a parede da sala com o cliente”, ou “Síndico reclamou que a equipe de obras deixa o elevador sujo ao fim do dia”.

Iniciou-se a análise com a contagem das visitas com anotações.

	1 <sup>2</sup> <sub>3</sub> visitas_com_ anotacao		.00 percentual_com_ anotacao
1	169	1	23.47

Há um total de 169 visitas com anotações no aplicativo, que representam 23,47% do total das visitas. Este dado não era esperado. Esperava-se que os usuários que tivessem engajados com o ato de criação de visita fizessem uso da parte de anotações. Talvez uma explicação seja o UI/UX não satisfatório para conduzir ao processo de anotações ou o fato de que os usuários fizessem sua comunicação com os clientes via whatsapp ou suas próprias anotações no bloco de notas do

celular ou mesmo em um caderno físico. Esta se mostrou uma feature que não valeria a pena ser mantida em versões futuras do Gelato.

Posteriormente analisou-se quais são os usuários que mais usam a funcionalidade das anotações, mas por motivo de privacidade estes dados são omitidos aqui, assim como o texto das anotações e em qual contexto elas se encaixavam. Esta análise não era possível sem o uso de SQL e foi uma experiência muito rica para entender o perfil das anotações dos usuários.

O que se pode mostrar aqui é uma distribuição com o número de caracteres das anotações que se mostraram bem distribuídas entre as categorias de número de caracteres até o limite de 300 caracteres. Apenas 6 anotações com mais de 600 caracteres.

	$A_C^B$ faixa	$1_3^2$ qtd_visitas
1	151–300 caracter...	33
2	1–50 caracteres	51
3	301–600 caracter...	18
4	51–150 caracteres	61
5	600+ caracteres	6

Há uma discussão interna no Gelato que ainda não foi resolvida sobre o custo x retorno de manter a funcionalidade das anotações. Por um lado, é muito rica a informação qualitativa que vem das tabelas, mas por outro é uma funcionalidade com relativamente pouco uso entre os usuários que consumiria mais recursos para ser implementada.

## AUTOAVALIAÇÃO E CONCLUSÃO

A experiência de ter criado um Data Warehouse e fazer as análises sobre os dados no Gelato usando SQL foi uma experiência riquíssima. Até então só era possível analisar os dados dentro das limitações do Excel e o uso do JOIN e LEFT JOIN para a construção de tabelas fato abriu um novo leque de possibilidades de análise.

Deve-se dizer que a execução do trabalho foi um exercício não trivial para a autora do trabalho dada a sua inexperiência na área de informática e a ambição dos objetivos traçados para as perguntas a serem respondidas. O fato de ter se escolhido um Modelo Estrela com 3 fatos e 6 dimensões trouxe inúmeras dificuldades para operacionalizar todo que se propôs a fazer, porém as conclusões que foram tiradas dos dados trouxeram uma recompensa substancial não só a nível pessoal, mas também para a empresa.

Um ponto que poderia ter sido melhorado no trabalho seria o reporte dos dados relativos aos usuários criando *alias* a eles de forma que seus e-mails ficassem ocultos e não houvesse problema com a exposição dos dados. Porém dada a complexidade do modelo adotado e restrições de tempo isso não foi explorado.

Poderia ter sido dedicado mais tempo para a exploração das palavras utilizadas nas anotações no app. Para isso deveria categorizar as palavras e fazer contagens dentro desta categoria. Esta é uma extensão deste trabalho que permitirá avaliar a função anotações com mais acurácia.

Outra melhoria futura que pode ser aplicada ao trabalho é a carga, transformação e análise dos dados de ambiente e serviço que não estavam salvos nas tabelas do banco de dados onde se esperava que eles estivessem.

Avalia-se que o trabalho feito possibilitou cumprir plenamente os objetivos de entender quais melhorias devem ser priorizadas na nova versão do aplicativo. Elas são:

- a) **Processo de onboarding intuitivo:** Hoje há interesse no aplicativo, as campanhas de publicidade conseguem captar leads de arquitetos para se inscreverem na plataforma, mas estes arquitetos não conseguem vencer as barreiras de criar seu primeiro projeto e sua primeira visita de forma fácil e intuitiva. Viu-se nos dados uma “desistência” em vencer etapas no processo de preenchimento do app.
- b) **Melhorar o upload de fotos:** os arquitetos não parecem estar usando o app na obra. Eles tiram as fotos em seus celulares e visitam a obra na terça-feira, mas só fazem o upload no

sistema sexta-feira. Um motivo para isso pode ser o processo lento de upload e a necessidade de ter que se usar wi-fi para isso. Caso o Gelato fosse um app nativo o usuário poderia fazer todo o processo sem precisar estar online.

- c) **Priorizar totalmente a experiencia relacionada a fotos:** as fotos estão presentes em 75% das visitas criadas e parecem ser a motivação principal do uso do Gelato. Portanto deve-se investir todos os recursos possíveis em tornar essa experiencia a mais rápida, eficiente e sem fricções possíveis. O caminho da melhora do engajamento no aplicativo está aqui.
- d) **Retirar funcionalidade das marcações:** Ela é um dos responsáveis pelo upload das fotos ser lento e é usada por poucos usuários. Seu ganho não vale o custo.
- e) **As pendências não são tão importantes quanto se imaginava:** Esperava-se que mais visitas tivessem pendências. Antes só era possível saber o volume total de pendências, mas não em quantas visitas diferentes essas pendências se distribuíam. Ao ver que menos de 25% das visitas criadas têm pendências, percebeu-se que elas não são críticas para as visitas, ao contrário das fotos.

A partir da análise feita neste trabalho e às conclusões que vieram dos dados reais de uso será possível criar uma segunda versão do aplicativo de forma mais otimizada e eficiente, aplicando os recursos escassos da empresa em funcionalidades que pode-se provar que geram valor aos usuários.

## ANEXO: CATÁLOGO DOS DADOS

### 1- Tabelas de Fatos

#### 1.1 – Fato Visita Criada

A tabela FATO\_VISITA\_CRIADA representa cada evento de criação de visita realizada pelos usuários dentro dos projetos cadastrados no Gelato.

Cada linha corresponde a uma visita única e contém as chaves estrangeiras que ligam o evento às dimensões relevantes (usuário, projeto, escritório e tempo), medidas agregáveis que permitem análises de engajamento (quantidade de fotos e pendências associadas) e atributos temporais e indicadores de modificação.

A granularidade da tabela é de uma linha por visita criada e sua representação pode ser vista abaixo.

Fato_Visita_Criada		
PK	id_fato_visita	INT (surrogate key)
FK	projeto_id	INT
FK	usuario_id	INT
FK	escritorio_id	INT
FK	tempo_id	INT
	nome_visita	STRING
	qtd_Fotos	INT
	qtd_Pendencias	INT
	foi_modificada	BOOLEAN
	date_vis	DATE
	date_vis_mod	DATE

Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição	Domínio
id_fato_visita	PK, INT (surrogate key)	Surrogate key gerada no DW para identificar cada fato de visita	Inteiro positivo, gerado automaticamente pelo DW, único e não nulo
projeto_id	FK para DIM_PROJETO	Projeto ao qual a visita pertence	Inteiro positivo
usuario_id	FK para DIM_USUARIO	Usuário que criou a visita	Inteiro positivo
escritorio_id	FK para DIM_ESCRITORIO	Escritório associado ao usuário criador	Inteiro positivo
tempo_id	FK para DIM_TEMPO	Data da visita	Inteiro positivo
nome_visita	STRING	Nome que o usuário deu à visita	Texto livre
qtd_Fotos	INT	Número total de fotos associadas à visita	Inteiro >= 0
qtd_Pendencias	INT	Número total de pendências criadas nesta visita	Inteiro >= 0
foi_modificada	BOOLEAN	Indica se a visita foi modificada (Modified Date ≠ Creation Date)	True ou False
DATE_vis	DATE	Data e hora da criação da visita (Creation date do Bubble)	Data válida
DATE_vis_mod	DATE	Data e hora da modificação da visita (Modified date do Bubble)	Data válida

## 1.2 – Fato Foto Adicionada

A tabela FATO\_FOTO\_ADICIONADA representa cada evento de adição de foto realizado durante uma visita técnica no Gelato.

Cada linha corresponde a uma foto individual, permitindo análises detalhadas de comportamento do usuário, distribuição de fotos por ambiente e serviço, existência e quantidade de tags e padrões de uso do aplicativo.

A tabela integra as principais dimensões envolvidas no registro da foto (projeto, usuário, escritório, tempo, visita) e inclui indicadores úteis para análises de qualidade e completude dos dados, como presença de marcador (como se fosse um comentário) na foto, uso posterior da foto (se ela é visível ao cliente ou não) e existência de tags de ambientes e serviços.

Aqui optou-se por simplificar o tratamento das tags que o usuário pode por na foto. Ao invés de criar novas dimensões com os ambientes e serviços existentes atribuídos à foto, mostra-se apenas se há tags usadas nas fotos, sua quantidade e armazena-se as tags para inspeção futura.

A granularidade da tabela é uma linha por foto adicionada e a representação da tabela pode ser vista abaixo.

Foto_Foto_Adicionada		
PK	id_fato_foto	INT (surrogate key)
FK	projeto_id	INT
FK	usuario_id	INT
FK	escritorio_id	INT
FK	tempo_id	INT
FK	visita_id	INT
	tem_tags	BOOLEAN
	qtd_tags	INT
	tags_raw	STRING
	foto_tem_marcador	BOOLEAN
	date_foto	DATE
	foto_cliente	BOOLEAN

Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição	Domínio dos Dados
id_fato_foto	PK, INT (surrogate key)	Surrogate key gerada no DW para identificar cada fato de foto	Inteiro positivo, gerado automaticamente pelo DW, único e não nulo
projeto_id	FK para DIM_PROJETO	Projeto ao qual a foto pertence	Inteiro positivo
usuario_id	FK para DIM_USUARIO	Usuário que adicionou a foto	Inteiro positivo
escritorio_id	FK para DIM_ESCRITORIO	Escritório associado ao usuário autor da foto	Inteiro positivo
tempo_id	FK para DIM_TEMPO	Data/hora em que a foto foi registrada	Inteiro positivo
visita_id	FK para FATO_VISITA_CRIADA	Visita na qual a foto foi adicionada	Inteiro positivo
tem_tags	BOOLEAN	Ambiente relacionado à foto, quando aplicável	True ou false
qtd_tags	INT	Serviço relacionado à foto, quando aplicável	Inteiro positivo

Tags_raw	STRING	Lista original das tags para inspeção manual	Texto livre
foto_tem_marcardor	BOOLEAN	Indica se a foto recebeu algum marcador (comentário) no app	True ou False
DATE_foto	DATE	Data e hora exata em que a foto foi registrada no sistema (Creation date do Bubble)	Data válida
foto_cliente	BOOLEAN	Indica se a foto é visível ao cliente (TRUE) não (FALSE)	True ou false

### 1.3 – Fato Pendência Criada

A tabela FATO\_PENDENCIA\_CRIADA representa cada evento de criação de pendência durante uma visita técnica ou processo de acompanhamento dentro dos projetos do Gelato. Cada linha corresponde a uma pendência individual com problemas identificados no imóvel a serem resolvidos.

A tabela incorpora todas as dimensões necessárias para contextualizar o evento (usuário, projeto, escritório, tempo, visita), além de um vínculo direto com a descrição da pendência — armazenada na dimensão correspondente. Ela também captura indicadores operacionais importantes como urgência, resolução, necessidade de envolver o cliente e datas de criação e modificação da pendência.

A granularidade desta tabela é uma linha por pendência criada e a representação da tabela pode ser vista abaixo:



Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição	Domínio dos Dados
id_fato_pendencia	PK, INT (surrogate key)	Identificador único gerado no DW para o fato de pendência	Inteiro positivo, gerado automaticamente pelo DW,

			único e não nulo
usuario_id	FK para DIM_USUARIO	Usuário responsável pela criação da pendência	Inteiro positivo
projeto_id	FK para DIM_PROJETO	Projeto ao qual a pendência pertence	Inteiro positivo
visita_id	FK para FATO_VISITA_CRIADA	Visita na qual a pendência foi identificada	Inteiro positivo
escritorio_id	FK para DIM_ESCRITORIO	Escritório associado ao usuário criador	Inteiro positivo
tempo_id	FK para DIM_TEMPO	Data da criação da pendência	Inteiro positivo
pendencia_desc_id	FK para DIM_PENDENCIA_DESC	Identificador da descrição textual da pendência	Inteiro positivo
eh_urgente	BOOLEAN	Indica se a pendência foi marcada como urgente no app	True ou False
resolvida	BOOLEAN	Se já foi resolvida (TRUE) ou não (FALSE)	True ou False
envolver_cliente	BOOLEAN	É necessária uma ação do cliente (TRUE) ou não (FALSE)	True ou False
DATE_pendencia	DATE	Data e hora da criação da visita (Creation date do Bubble)	Data válida
DATE_pendencia_mod	DATE	Data e hora da modificação da visita (Modified date do Bubble)	Data válida

## 1.4 – Dimensão Usuário

A DIM\_USUARIO armazena os atributos descritivos dos usuários que interagem com o Gelato.

Esta dimensão permite análises relacionadas ao perfil do usuário, ao escritório ao qual ele está vinculado, à sua assinatura e seus dados de contato (como e-mail e telefone). Vale ressaltar que o identificador único do usuário dentro do Bubble é o e-mail do usuário, então não foi necessário criar um campo extra para a identificação do usuário no OLTP.

Cada linha corresponde a um usuário único, contendo informações cadastrais, características de uso e vínculos organizacionais.

A granularidade da dimensão é uma linha por usuário.

dim_Usuario		
PK	usuario_id	INT (surrogate key)
	nome	STRING
	email	STRING
	plano	STRING
	telefone	STRING
	data_criacao	DATE
	escritorio_nome	STRING

Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição	Domínio dos Dados
usuario_id	PK, INT (surrogate key)	Identificador único do usuário dentro do DW	Inteiro positivo, gerado automaticamente pelo DW,



			único e não nulo
nome	STRING	Nome completo do usuário	Texto livre
email	STRING	E-mail utilizado para login e identificação do usuário no OLTP	Texto livre
plano	STRING	Tipo de plano do usuário no app (Grátis, Intermediário, Avançado)	Texto da lista de planos
telefone	STRING	Telefone de contato informado no cadastro	Texto numérico até 20 caract podendo haver especiais
data_criacao	DATE	Data e hora em que o usuário foi criado no sistema (Creation date do Bubble)	Data valida
escritorio_nome	STRING	Nome do escritório associado ao usuário	Texto livre

## 1.5 – Dimensão Projeto

A dimensão dim\_Projeto armazena os atributos descritivos dos projetos cadastrados pelos usuários no Gelato.

Cada linha representa um projeto único e contém informações sobre o cliente, metragem, data de criação e volume de visitas previstas.

A granularidade desta dimensão é uma linha por projeto existente no sistema.



Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição	Domínio dos Dados
projeto_id	PK, INT (surrogate key)	Identificador único do projeto no DW	Inteiro positivo, gerado automaticamente pelo DW, único e não nulo
projeto_uuid	STRING	Identificador original do projeto no sistema transacional (UUID do Bubble)	Texto alfanumérico no formato UUID, valor único por projeto
nome_projeto	STRING	Nome atribuído ao projeto pelo usuário	Texto livre
nome_cliente	STRING	Nome do cliente/proprietário associado ao projeto	Texto livre
data_criacao	DATE	Data e hora de criação do projeto no sistema (Creation date do Bubble)	Data valida
metragem	FLOAT	Área total informada para o projeto, em metros quadrados	Numero real positivo
nro_visitas_previstas	INT	Número de visitas que o usuário espera que o projeto tenha	Inteiro >=0

## 1.6 – Dimensão Tempo

A dimensão dim\_Tempo armazena atributos derivados do calendário e é utilizada para relacionar eventos do sistema a datas específicas.

Esta dimensão permite análises temporais como distribuição de eventos ao longo do mês, sazonalidade, comportamento por trimestre e dia da semana.

A granularidade da dimensão é uma linha por dia no calendário.

dim_Tempo		
PK	tempo_id	INT (surrogate key)
	data	DATE
	dia	INT
	mes	INT
	ano	INT
	trimestre	INT
	dia_semana	STRING

Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição	Domínio dos Dados
tempo_id	PK, INT (surrogate key)	Identificador único da data no DW	Inteiro positivo, gerado automaticamente pelo DW, único e não nulo
data	DATE	Data completa no formato AAAA-MM-DD	Data valida
dia	INT	Número do dia do mês	Inteiro 1-31
mes	INT	Número do mês (1 a 12)	Inteiro 1-12
ano	INT	Ano calendário	inteiro com 4 dígitos >= 1900
trimestre	INT	Trimestre do ano (1 a 4)	1, 2, 3, 4
dia_semana	STRING	Nome do dia da semana correspondente à data	Texto dias da semana

## 1.7 – Dimensão Visita

A dimensão dim\_Visita armazena os atributos descritivos das visitas registradas pelos usuários dentro dos projetos do Gelato.

Cada linha representa uma visita única, permitindo ver a nomenclatura utilizada e as anotações feitas no sistema.

A granularidade desta dimensão é uma linha por visita criada.

dim_Visita		
PK	visita_id	INT (surrogate key)
	visita_uuid	STRING
	nome_visita	STRING
	anotacoes	STRING

Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição	Domínio dos Dados
------	------	-----------	-------------------

visita_id	PK, INT (surrogate key)	Identificador único da visita no DW	Inteiro positivo, gerado automaticamente pelo DW, único e não nulo
visita_uuid	STRING	Identificador original da visita no OLTP (UUID do Bubble)	Texto alfanumérico no formato UUID, valor único por projeto
nome_visita	STRING	Nome atribuído à visita pelo usuário	Texto livre
anotacoes	STRING	Texto livre contendo anotações registradas na visita	Texto livre

## 1.8 – Dimensão Pendência

A dimensão dim\_Pendencia armazena exclusivamente o texto descritivo das pendências criadas pelos usuários no sistema.

Como os detalhes operacionais, indicadores e vínculos da pendência já estão representados na tabela de fatos FATO\_PENDENCIA\_CRIADA, esta dimensão existe apenas para normalizar e padronizar o campo textual, evitando redundância e permitindo análises mais organizadas sobre o conteúdo das pendências registradas.

A granularidade desta dimensão é uma linha por texto de pendência único.

dim_Pendencia			
PK	pendencia_id	INT (surrogate key)	
	pendencia_uuid	STRING	
	texto_pendencia	STRING	

Os campos da tabela podem ser descritos da seguinte forma:

Nome	Tipo	Descrição
pendencia_id	PK, INT (surrogate key)	Identificador único da descrição no DW
pendencia_uuid	STRING	Identificador original da pendência no OLTP
texto_pendencia	STRING	Texto descritivo da pendência registrada pelo usuário