Mineração de Dados e Uso de Dashboards para Visualização de Dados e Suporte à Tomada de Decisões

Data Mining and Dashboards Use for Data Visualization and Decision Making Support

Eduardo Correia ¹ **orcid.org/0000-0002-8043-6087**

Francisco Guimarães ¹ <u>orcid.org/</u>

Ruben Araujo ¹ **orcid.org/0000-0002-7633-1581**

E-mails dos autores principais: Eduardo Correia evssc@ecomp.poli.br; **Francisco Guimarães** fvcg@ecomp.poli.br; **Lucas Leite** llc@ecomp.poli.br; **Ruben Araujo** rfga@ecomp.poli.br

Resumo

Dentre os serviços virtuais disponibilizados por algumas empresas, está o monitoramento de sistemas computacionais para mitigação de erros ou falhas nos serviços de um cliente. Em muitos casos, uma estrutura de *Service Desk* é montada para prover a manutenção desses sistemas, sob um acordo de nível de serviço firmado entre o cliente e a prestadora do serviço de monitoramento. Para assegurar efetivos acompanhamento e respostas a eventuais falhas desses sistemas (em alguns casos chamados de *jobs*), um processo de Mineração de Dados se mostra altamente promissor para ampliar a visão sobre o serviço prestado e, consequentemente, promover *insights* que permitam uma constante otimização dos processos envolvidos, tendo por base os dados gerados diariamente a partir do monitoramento das atividades dos sistemas de clientes. Assim, este artigo propõe uma análise detalhada de uma base de dados sobre execução de *jobs*, seguindo a metodologia CRISP-DM. Os dados foram submetidos a experimentos de mineração e os resultados foram peça essencial para a construção de *dashboards*, os quais, através da ferramenta gratuita e online, Google Data Studio, servem para auxiliar na análise e tomada de decisão sobre o contexto do serviço de monitoramento provido.

Palavras-Chave: Dados; Erros; Processamento; Monitoramento; dashboards.

Abstract

Among the virtual services provided by some companies, there is the monitoring of computer systems to mitigate errors or failures in the services of a customer. In many cases, a Service Desk structure is set up to provide maintenance for these systems, under a service level agreement between the customer and the monitoring service provider. In order to ensure effective monitoring and responses to eventual failures of these systems (in some cases called jobs), a Data Mining process is highly promising to expand the view on the service provided and, consequently, promote insights that allow a constant optimization of processes involved, based on the data generated daily from monitoring the activities of customer systems. Thus, this article proposes a detailed analysis of a database about jobs' execution, following the CRISP-DM methodology. The data were subjected to mining experiments and the results were an essential part for the construction of dashboards, which, through the free and online tool, Google Data Studio, serve to assist in the analysis and decision making about the context of the monitoring service provided.

Key-words: Data; Errors; Processing; Monitoring, dashboards.

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil.

1 Introdução

1.1 Contextualização

Diante da imensa quantidade de serviços oferecidos de maneira virtual através da internet, é o caminho natural pensar continuamente em aspectos desses serviços que podem ser melhorados. A qualidade do atendimento e recursos fornecidos aos clientes é um desses aspectos.

Fatores como facilidade, tempo otimizado e custos reduzidos, são tópicos que atraem usuários para o uso desses serviços que estão se espalhando e ganhando espaço rapidamente pelos mais diversos âmbitos comerciais.

Assim, a necessidade de uma constante melhoria na eficácia do suporte oferecido aos consumidores, sejam estes empresas ou pessoas físicas, torna-se perceptível; uma vez que no mundo cada vez mais conectado, as expectativas dos clientes em relação à qualidade desses serviços tendem a crescer.

De acordo com o relatório State of Customer Service da Microsoft [1], 76% de clientes brasileiros já desistiram de realizar negócios com uma empresa por causa de má experiência com o atendimento. A pesquisa, realizada em 2018, contou com a participação de 5000 indivíduos em cinco países – Alemanha, Brasil, Japão, Reino Unido e Estados Unidos. Seus resultados confirmam que o serviço ao consumidor é um dos pilares da experiência do cliente, servindo também como diferencial para as empresas que pretendem destacar-se no mercado.

De acordo com a mesma pesquisa, 95% dos participantes indicaram que o serviço ao consumidor é importante para suas escolhas em relação à lealdade a um prestador de serviço ou marca.

Ao entender o presente cenário, torna-se notável a relevância que um atendimento eficiente e preciso faz, não só pela satisfação do cliente, mas também pela imagem da empresa oferecendo seu serviço.

1.2 Descrição do Problema

Dentre os serviços virtuais disponibilizados por algumas empresas, está o monitoramento de sistemas computacionais para mitigação de falhas nos serviços de um cliente. No processo de contratação desses serviços, é definido um SLA (Service Level Agrement) ou Acordo de Nível de Serviço, através do qual serão firmados, dentro vários outros fatores, metas de desempenho quanto ao tempo e a qualidade da resposta ao uma determinada falha, para que a prestação do serviço seja tida como satisfatória.

Diante desse cenário, um atendimento proativo ou reativo poderá ser disponibilizado pelo provedor. No atendimento proativo, a central de monitoramento do serviço iniciará a tratativa do possível problema antes que o consumidor sinta os efeitos e inicie um pedido de suporte técnico. No atendimento reativo, o cliente, seguindo prévia orientação, iniciará o relato de um problema através de uma central única de atendimento e, a partir daí, a solicitação encaminhada para dos analistas um disponíveis, que dará a devida tratativa para a solução do problema.

Porém, sempre apresenta-se simples; sem suas causas raízes claramente definidas, o problema se torna difícil de explorar e uma visão analítica deste é praticamente inviável. Seja pela falta de organização quanto aos seus processos ou análises inadequadas das informações disponíveis, o provedor do serviço pode não oferecer uma visão precisa dos resultados da operação na resolução dos chamados aos seus clientes, nem tão pouco prover um bom atendimento proativo.

1.3 Objetivo

O objetivo deste trabalho leva em consideração o cenário no qual há o relacionamento cliente-empresa, em que a empresa é responsável por gerenciar e monitorar

ambientes virtuais – sistemas – dos seus clientes, rastreando o processamento de dados nos sistemas ou o funcionamento de alguma área específica destes.

Assim, almeja-se o desenvolvimento de um método ou modelo de análise que aponte os principais causadores de problemas que possam vir a ocorrer nesses sistemas – a partir da investigação de dados gerados pelo funcionamento dos processos internos e coletados nesses ambientes.

Pretende-se também que seja possível construir um panorama abrangente – com base na análise realizada – que auxilie na elaboração de estratégias para melhor alocação e entendimento dos processos, de maneira que medidas possam ser impostas antes de problemas ocorrerem.

1.4 Justificativa

Um método de visualização e análise mais preciso permite que a equipe de analistas, responsável por tratar os problemas dos clientes, tenha uma visão interna mais exata de seus sistemas, e, assim, possa oferecer o suporte mais adequado às necessidades de seus clientes. Além disso, observa-se o trabalho de maneira minuciosa, obtém-se a chance de melhor acompanhamento da própria performance; investigando problemas realmente relevantes e gerando relatórios mais factíveis sobre o desempenho atual do monitoramento.

1.5 Escopo Negativo

Este artigo não visa esgotar as possibilidades de tratamento e extração de informações da base de dados. Nem objetiva a integração com bases de dados externas além da estudada.

2 Fundamentação Teórica

2.1 Service Desk

De acordo com o glossário de termos técnicos da Gartner [2], Service Desk "é um centro de ajuda que está equipado com os recursos para a resolução de solicitações de serviço e problemas. Ele dá ao prestador de

serviço e ao cliente (usuário final) a capacidade de diagnosticar de forma eficiente e corrigir problemas de suporte técnico".

Uma forma de ver o Service Desk é como uma evolução do Help Desk – um serviço de atendimento ao cliente que centraliza chamados em uma única plataforma, funcionando como um suporte técnico.

A diferença entre ambos é que o Help Desk normalmente resolve questões técnicas mais simples, como a falta de acesso à Internet ou a dificuldade de manusear algum software, enquanto o Service Desk vai mais distante.

Além de oferecer o tipo de auxílio que o Help Desk oferece, o Service Desk funciona também como uma espécie de consultoria na área de TI, de forma que os clientes possam contar com a ajuda de especialistas do setor através de propostas de melhoria técnica e acompanhamento.

O Service Desk concentra muito mais demandas que o seu "antecessor" e centraliza todas as necessidades da empresa em relação à tecnologia da informação em um só lugar, oferecendo não só um suporte técnico na área, mas também um auxílio estratégico.

Enquanto o Help Desk pode ser associado a um "primeiro nível de atendimento", o Service Desk ganha o título de "Segundo Nível de Atendimento" pelo fato de envolver demandas mais complexas.

Pelo motivo de ser necessária a realização de um trabalho mais complexo, a equipe de Service Desk costuma ser mais capacitada tecnicamente que a de Help Desk. Logo, os membros da equipe de analistas precisam ter um amplo conhecimento técnico e dominar os processos envolvidos com as práticas de TI.

Assim, o Service Desk possui a responsabilidade de promover a manutenção de TI da organização, evitando panes gerais que podem causar prejuízos.

Ele também serve para acompanhar os ciclos de vida dos pacotes de serviços, fornecendo um fluxo de informações críticas para uma melhoria contínua – principalmente no que diz respeito à gestão dos atendimentos.

Destacam-se como vantagens do Service Desk:

- Ajuda nas estratégias da empresa;
- Auxilia no crescimento programado da organização;
- Potencializa o suporte de TI;

• Tem caráter consultivo;

O modelo de um Service Desk pode ser estruturado de três formas:

- Service Desk local: quando os usuários e profissionais da central de serviços estão próximos (no mesmo prédio, geralmente);
- Service Desk centralizado: quando os usuários de diferentes locais precisam fazer suas requisições por meio de uma central, que está fisicamente localizada em um ponto único;
- Service Desk Virtual: sem uma localização física definida, pode estar em diferentes locais (em vários países, por exemplo). O atendimento aos usuários é feito unicamente via web.

2.2 Mineração de Dados

A mineração de dados surgiu no atual contexto de superabundância de dados – gerados a partir do processo de avanço da tecnologia e da interação cada vez maior das pessoas com o mundo digital – "como um processo sistemático, interativo e iterativo, de preparação e extração de conhecimentos a partir de grandes bases de dados" (CASTRO, 2016, p.44).

O termo é usado como alusão ao processo de extração de minerais valiosos a partir de uma mina, de maneira que no contexto dos dados a mina seria a base de dados, os algoritmos utilizados, as ferramentas utilizadas para a extração, e o conhecimento extraído, os minerais preciosos.

Os dados são símbolos ou signos como valores em uma tabela, e a informação está contida nas descrições desses símbolos – a exemplo dos títulos das colunas em uma planilha. E o conhecimento gerado a partir da análise desses dados é algo que permite a tomada de decisões para a agregação de valor em determinado âmbito.

A mineração de dados é uma disciplina interdisciplinar e multidisciplinar que envolve conhecimentos de áreas como banco de dados, estatística, aprendizagem de máquina,

visualização de dados, inteligência artificial, entre outras; e é parte de um processo mais amplo, conhecido como descoberta de conhecimento em bases de dados (knowledge Discovery in databases, KDD), geralmente sintetizado nos seguintes tópicos:

- Base de dados;
- Preparação ou pré-processamento de dados;
- Mineração de dados;
- Avaliação ou validação do conhecimento.

O pré-processamento dos dados é uma das etapas mais importantes da mineração de dados. Ela é nada mais que o processo de preparação, organização e estruturação dos dados. Podemos destacar 3 principais passos do pré-processamento:

- A limpeza dos dados, que envolve o manuseio ou preenchimento de dados faltantes, identificação e remoção de valores aberrantes, redução de ruídos e a resolução de inconsistências.
- A transformação dos dados é a etapa em que os dados serão formatados para que se adequem ao processo de mineração. Envolve a normalização, seleção e geração de atributos.
- 3. A redução dos dados é útil para diminuir o processamento computacional necessário para a mineração e aumentar a eficiência.

2.2.1 **CRISP-DM**

Cross Industry Standard Process for Data Mining (Processo Padrão Entre Setores para Mineração de Dados, em tradução livre), esta é a definição de CRISP-DM. Este é um modelo de processo, concebido em 1996 [4] e composto por 6 (seis) fases que descrevem o ciclo de vida de um projeto em ciência de dados alinhado com o negócio (business) [3].

A imagem seguinte representa esse processo. Nela podemos observar que se trata de

um processo com muita interação entre as fases. Além disso, é iterativo, com as ações de uma fase dependendo das anteriores, as quais precisam estar bem desenvolvidas para o sucesso das posteriores.

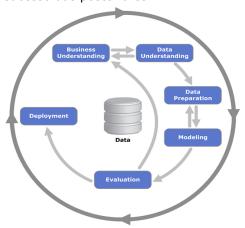


Figura 1: Fluxo do CRISP-DM. Fonte: Wikimedia Commons (2012).

Como é possível observar, esse processo engloba as seguintes etapas:

- Entendimento (das necessidades) do negócio;
- 2. Entendimentos dos dados;
- 3. Preparação dos dados (para facilitação das análises subsequentes);
- Modelagem: aplicação de técnicas de mineração de dados sobre os dados disponíveis;
- 5. Avaliação: avaliação dos resultados obtidos na fase anterior;
- Implementação: disponibilização dos resultados úteis como ferramenta para a otimização do negócio.

2.2.2 Uso de *Dashboards* na Visualização de Dados

Dashboards são painéis visuais que apresentam um conjunto de informações - como métricas e indicadores de desempenho - acerca dos processos de uma empresa, serviços, sistemas e afins de maneira centralizada. O principal objetivo dessas ferramentas é fornecer, de maneira acessível e facilitada, a possibilidade do acompanhamento dos processos intrínsecos a um determinado ecossistema que produz dados como resposta a suas atividades.

O uso de dashboards apela para o sentido visual das pessoas, tornando o uso da ferramenta de extrema utilidade para análises de eventos e tomadas de decisões mais rápidas, além de ser útil para o suporte à definição de metas e objetivos.

Os dashboards possuem a capacidade de promover, juntamente com suas métricas e indicadores, o monitoramento do desempenho de atividades, de forma a tornar mais cômoda, para seus observadores, a leitura dos dados e a consequente inferência de informações.

Eles possibilitam a personalização de painéis com a mostra de dados específicos e a filtragem de informações isoladas ou em conjunto para uma melhor visualização do que está acontecendo nos processos.

Uma vantagem adicional do uso de dashboards é a capacidade de lidar com excesso de informação proveniente de uma grande quantidade de dados gerados a cada instante; característica de vários sistemas atualmente. Esse excesso pode obscurecer os processos de análise e dificultar a tomada de decisões.

Assim, por meio dos dashboards essas informações podem ser disponibilizadas de uma maneira mais sintetizada e inteligível para os usuários em situações operacionais, táticas e/ou estratégicas.

3 Materiais e Métodos

O ciclo de desenvolvimento do projeto teve como base o CRISP-DM. Assim, os materiais obtidos foram analisados para se chegar aos resultados atuais. Tendo já entendido o contexto de aplicação do projeto, partimos para o entendimento, preparação e aplicação de modelagens sobre a base de dados obtida.

3.1 Descrição da Base de Dados

A base de dados, já estruturada, utilizada para a análise do problema, contém dados provenientes do monitoramento feito por um algoritmo executado automaticamente no ambiente de software de clientes. Esse algoritmo foi denominado por robô P2D2 e inspirou também o nome da base de dados, a qual disponibiliza 18735 registros do período de 10/11/2018 a 09/01/2019. As informações presentes nessa base são:

Nome do cliente;

- Nome do job;
- Ambiente no qual o job está sendo/foi executado;
- Data e horário de execução do job;
- Tempo de execução do job;
- Meta (tempo esperado) de execução do job;
- Status do job.

Além de informações acerca da escala de trabalho dos analistas atendentes.

Tendo em vista essas informações, também é importante esclarecer que a variação dos dados, em parte, foi originalmente gerada artificialmente para este projeto.

3.2 Dicionário de Dados

A Tabela 1 apresenta o dicionário de dados da base de dados utilizada. A tabela descreve os dados encontrados na base quanto ao tipo e suas finalidades.

Tabela 1: Dicionário de dados da base de dados.

Coluna	Tipo	Valores Possíveis	Descrição
NOME_DO_CLIENTE	Texto		Identificador do cliente no sistema.
AMBIENTE	Texto		Identificador do ambiente no sistema.
JOB	Texto		Identificador do processo no sistema.
DATA_EXECUCAO	Data		Data na qual a execução listada do processo foi realizada.
HORA_INICIO	Temporal		Hora na qual a execução listada do processo foi iniciada.
HORA_FIM	Temporal		Hora na qual a execução listada do processo foi finalizada.
TEMPO_EXECUCAO	Numérico		Tempo (em segundos) total da execução do processo.
META_ATUAL_DE_EXECUCAO	Numérico		Tempo máximo (em segundos) no qual o processo é esperado ser executado.
STATUS	Texto	PLANEJADO, EM ANDAMENTO, OK, ERRO	Estado do processo durante a leitura do servidor para coleta dos dados.
TEMPO_RESTANTE	Numérico		Resultante da fase de pré-processamento. Apresenta o cumprimento ou não da meta atual de execução: TEMPO_EXECUCAO subtraído de META_ATUAL_DE_EXECUCAO
HORA_INICIO_NUM	Numérico		Resultante da fase de pré-processamento. Representa apenas o valor da hora de início do processamento do <i>job</i> em questão (sem minutos e segundos)
HORA_FIM_NUM	Numérico		Resultante da fase de pré-processamento. Representa apenas o valor da hora de término do processamento do job em questão (sem minutos e segundos)

3.3 Análise Descritiva dos Dados

Esta etapa foi de grandíssima relevância para a geração de valor das análises. Ela é altamente evidenciada até mesmo na fase de *deployment* do projeto descrito neste artigo.

Isto, porque vários algoritmos foram testados (regressão, agrupamento, classificação, etc.) sobre a base de dados, na busca por indicadores que promovessem conclusões assertivas para a tomada de decisão quanto a melhoria no processo de trabalho no campo de atuação já mencionado. Contudo, os resultados obtidos com a análise descritiva dos dados se mostraram mais relevantes e confiáveis, dentre outros fatores,

devido a baixa incidência de padrões nas variações dos dados e o tempo disponível para otimizar os experimentos.

Na sessão seguinte serão apresentados alguns resultados dessa análise, com tabelas produzidas pelos autores.

3.3.1 Distribuição de Frequência

Tabela 2: Distribuição da frequência absoluta de *jobs* executados com erro por ambiente.

Ambiente	Erros
AMB9	651
AMB2	313
AMB1	311

	AMB3	274
	AMB4	258
	AMB5	224
	AMB7	26
	AMB8	19
	AMB8	131
•	AMB6	15

Tabela 3: Distribuição da frequência absoluta de *jobs* bem sucedidos por ambiente.

Ambiente	ОК
AMB9	4033
AMB1	1790
AMB3	1752
AMB2	1717
AMB4	1501
AMB5	1381
AMB7	108
AMB6	96
AMB8	83
AMB10	1

Tabela 4: Distribuição da frequência absoluta de todos os *jobs* por ambiente.

Ambiente	Total
AMB9	4684
AMB1	2101
AMB2	2030
AMB3	2026
AMB4	1759
AMB5	1605
AMB7	134
AMB6	111
AMB8	102
AMB10	1

Tabela 5: Distribuição da frequência relativa de *jobs* executados com erro por ambiente.

Ambiente	Frequência Relativa
AMB9	31,13%
AMB2	14,97%
AMB1	14,87%
AMB3	13,10%
AMB4	12,34%
AMB5	10,71%
AMB7	1,24%
AMB8	0,91%
AMB6	0,72%

Tabela 6: Distribuição da frequência relativa de *jobs* bem sucedidos por ambiente.

Ambiente	Frequência Relativa
AMB9	32,36%
AMB1	14,36%
AMB3	14,06%
AMB2	13,8%
AMB4	12,04%
AMB5	11,08%
AMB7	0,86%
AMB6	0,77%
AMB8	0,66%
AMB10	0,008%

Tabela 7: Distribuição da frequência relativa de todos os *jobs* por ambiente.

Ambiente	Frequência Relativa
AMB9	32,19%
AMB1	14,44%
AMB2	13,95%
AMB3	13,92%
AMB4	12,09%
AMB5	11,03%
AMB7	0,92%
AMB6	0,76%
AMB8	0,70%
AMB10	0,007%

3.3.2 Visualização dos Dados

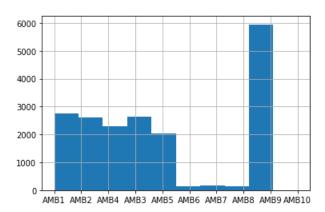


Figura 1: Histograma relacionando os ambientes à quantidade de *jobs* existentes.

Fonte: os autores.

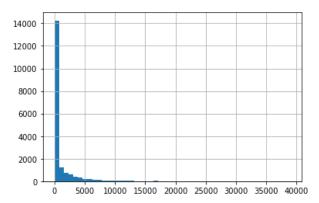


Figura 2: Histograma relacionando o tempo de execução à quantidade de *jobs*.

Fonte: os autores.

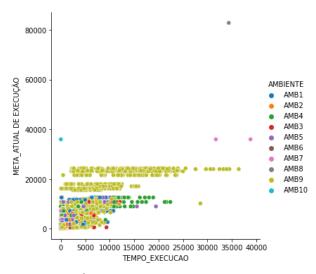


Figura 3: Gráfico de dispersão relacionando o tempo de execução à meta atual de execução por ambiente. Fonte: os autores.

3.4 Pré-processamento dos Dados

Durante esta fase do projeto, algumas modificações foram necessárias para o aprimoramento da base de dados. As alterações a seguir, incluindo também a readaptação do nome de alguns atributos para melhor manipulação em um modelo e já foram atualizadas na Tabela 1.

3.4.1 Limpeza de dados

As entradas que possuíam o atributo STATUS como "Em andamento" ou "Planejado" foram

removidas. Isto porque, por uma análise visual, ao observar a hora de finalização com a data de captura do *status* de alguns dos serviços do cliente, não se pôde obter uma informação conclusiva sobre o real processo de atividade desses serviços.

Contudo, posteriormente, essas entradas poderão ser incluídas na análise, tanto como um cluster separado das demais, quanto em experimentos incluindo todas.

3.4.2 Transformação de Dados

Neste processo, o atributo TEMPO_RESTANTE foi gerado, sendo resultado da operação de diferença entre META_ATUAL_DE_EXECUCAO e TEMPO_EXECUCAO. Esse importante atributo poderá auxiliar na análise de padrões de cumprimento de metas, relacionados - também - aos serviços e à escala dos analistas. Ele traz informações de infração (valores negativos) ou não (valores positivos) sobre o SLA acordado e também expõe alguma economia de tempo de execução.

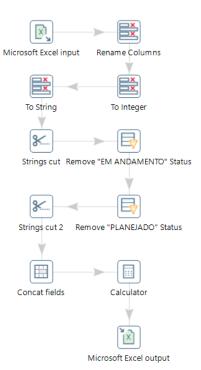


Figura 5: Passo a passo da transformação dos dados na ferramenta Pentaho.

Fonte: os autores.

3.5 Metodologia Experimental

Como já mencionado, o experimento que agregou mais valor ao projeto foi baseado na análise descritiva dos dados. Nas sessões subsequentes, será possível observar como os benefícios dela foram ampliados e postos em evidência.

3.5.1 Google Data Studio

Para realizar a análise dos dados após o processo de pré-processamento, foi utilizada a ferramenta Data Studio desenvolvida e oferecida, em versão gratuita, pela Google.

O Data Studio é uma ferramenta que possibilita a criação de relatórios e dashboards fáceis de ler, personalizar, e serem compartilhados juntamente com os dados, cujas fontes podem ser diversas e contínuas.

Exemplos de fontes de dados às quais o Data Studio pode ser conectado são: produtos do Google Marketing Platform, como o Google Ads e Analytics; banco de dados como o BigQuery, MySQL e PostgreSQL e arquivos simples como os do tipo CSV.

A partir da inserção na ferramenta, os dados podem ser visualizados através da perspectiva de vários tipos de gráficos, como os de linhas, *pizza*, barras, mapas geográficos e tabelas dinâmicas.

A criação de relatórios e dashboards interativos se torna possível por meio dos filtros de visualização e controles de períodos. Esse controle de dados torna os relatórios em modelos flexíveis com a possibilidade de portabilidade.

4 Análise e Discussão dos Resultados

Durante a aplicação de algoritmos sobre a base de dados, percebeu-se uma baixa incidência de padrões nas variações dos dados, o que possivelmente foi causado pela geração artificial de dados na base de dados original.

4.1 Resultados

A seguir estão dispostos alguns dos dashboards gerados com a análise descritiva dos dados pré-processados utilizados.



Figura 6: Perspectiva geral dos dados após o pré-processamento.

Fonte: os autores.

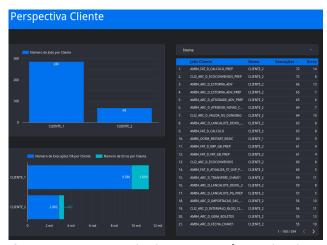


Figura 7: Perspectiva de clientes com número de *jobs* e seus status por cliente.

Fonte: os autores.



Figura 8: Perspectiva de tempo com porcentagem de atraso em relação à meta de execução.

Fonte: os autores.

5 Referências

- [1] State of Global Customer Service Report. Disponível em: https://info.microsoft.com/rs/157-GQE-382/images/EN-US-CNTNT-ebook-2018-State-of-Global-Customer-Service.pdf. Acesso em: 21 mar. 2021, 12:30:39.
- [2] Gartner Glossary, Information Technology Glossary. Disponível em: https://www.gartner.com/en/information-technology/glossary/service-desk. Acesso em: 21 mar. 2021, 13:10:15.
- [3] Data Science Project Management. Disponível em: https://www.datascience-pm.com/crisp-dm-2/ >. Acesso em: 12 mai. 2021, 14:00:00.
- [4] Cross-industry standard process for data mining. Disponível em: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining. Acesso em: 13 mai. 2021, 01:18:00.