

PROJETO FINAL – ANÁLISE DE DADOS TAURUS

EQUIPE ROVER SOLUTIONS

Professora: Franciane Rodrigues

Professor: Douglas Ribeiro

Professor: Jonathas Carneiro

## Sumário

[Sumário 1](#_djoqahcv1gmu)

[**Projeto Taurus - Rover Solutions 2**](#_ddyo48wkl0d3)

[Identificação 3](#_yo9ymxz7edmf)

[Critérios de Aceitação 3](#_unwvpt6nqx4m)

[Introdução 4](#_qdkqc2n5or63)

[Metodologia SCRUM 5](#_cwxv48oaao00)

[Datasets da Taurus, sua Origem 7](#_s84gran8r9r3)

[KPI’s Definidos 8](#_3cg2e35j0eue)

[Metodologia KDD 9](#_mtlnnzopkw5o)

[Workflow 9](#_x0ymoobkp44t)

[Análise de dados: Etapa Pandas 10](#_adj9e3j0eifp)

[Dataframe DRE: 10](#_sg8f1vuxyq6n)

[Dataframe Balanço Patrimonial: 14](#_62k6r69ed7z5)

[Análise de dados: Etapa Pyspark 17](#_i5f1r625q4d9)

[Dataframe DRE: 17](#_4vtewu7ki8le)

[Análise descritiva DRE 18](#_b3qfcgilztgd)

[Análise KPI’s DRE 21](#_2qudsgw7glqc)

[Análise Comparativa de receitas brutas dos trimestres de 2019,2020 e 2021 23](#_9cqz6y2xlce8)

[Dataframe Balanço Patrimonial: 25](#_b6v0g3v1gc4i)

[Análise descritiva DRE 26](#_70npk6ygl3xi)

[Análise KPI’s Balanço Patrimonial 29](#_erplbqjcq1c1)

[Google Cloud, MySQL e Bigquery 31](#_f7rg2ioz1zab)

[MongoDB 34](#_gvzrzihhh2ie)

[Análise Swot 36](#_p76irdmc7fci)

[Looker Studio 37](#_k1r3z3dtsb3h)

[Links Importantes 38](#_duzf7rafdc2h)

[Conclusão 38](#_sdazc73tvxx)

# Projeto Taurus - Rover Solutions

| 1. Informações Gerais do Projeto | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Empresa definida** | | | Taurus | | |
| **Premissa** | | | Qual é o impacto da variação dos preços de matéria-prima/insumos nos ganhos de rentabilidade? | | |
| **Mapeamento de Riscos** | | | Conter viés de parcialidade política, pessoal ou profissional; Fuga do escopo inicial. | | |
| **Restrições** | | | O projeto deverá ser entregue no prazo definido e acompanhado com base no cronograma scrum. | | |
| Todas as etapas deverão ser seguidas e revisadas dentro do prazo de entrega, minimizando qualquer falha futura. | | |
| **Scrum** | | | Team - 01 Finances/Accountabilities - Sprint 2 Weeks | | |
| 2. Equipe | | | | | |
|  | | Nome | | Email | |
| Líder da Equipe Rover | | Leandro de Oliveira | | profic.m3@gmail.com | |
| Membros da Equipe Rover | | Gabriel Henrique Moraes | | gabrielnutsu@gmail.com | |
| Ériv Jhomic | | az.jhomic@gmail.com | |
| Jonatan Dourado | | jnatansd@gmail.com | |
| Renata Reis | | natinha.amr@gmail.com | |
| Victor Gaspar | | viictor.gaspar@gmail.com | |

### **Identificação**

**Projeto**: Qual é o impacto da variação dos preços de matéria-prima/insumos nos ganhos de rentabilidade?

**Líder**: Leandro de Oliveira

**Orientador**: Douglas Ribeiro

**Patrocinador**: SoulCode Academy

### **Critérios de Aceitação**

1.0. APRESENTAÇÃO;

1.1. Deverá apresentar as funções e ferramentas utilizadas no código.

1.2. Explicar o porquê do dataset escolhido.

1.3. Todos os componentes (integrantes) deverão se apresentar

1.4. Deverá ser usado termos técnicos, evitando o uso de gírias ou

expressões coloquiais e/ou culturais.

1.5. Cada grupo terá até 60 minutos para se apresentar.

1.6. A ordem da apresentação será comunicada pelos professores próximo à

data de apresentação.

2.0.REQUISITOS OBRIGATÓRIOS;

2.1. Obrigatoriamente os datasets devem ter formatos diferentes (CSV / Json

/ Sql / NoSql/ Excel, outros) e 1 deles obrigatoriamente tem que ser em

CSV.

2.2. Operações com Pandas (limpezas , transformações e normatizações)

2.3. Operações de análise de dados usando PySpark com a justificativa de

cada uma das transformações utilizadas com o mesmo.

2.4. Utilizar pelo menos 2 gráficos distintos (bibliotecas python) para

representar os dados e encontrar possíveis inconsistências

2.5. Os datasets utilizados podem ser em língua estrangeira , mas devem ao

final terem seus dados/colunas exibidos na língua PT-BR (Utilizar a

diretrizes abordadas em aula)

2.6. Os datasets devem ser salvos e operados em armazenamento cloud

obrigatoriamente dentro da plataforma GCP (não pode ser usado Google

drive ou armazenamento alheio ao google)

2.7. Os dados tratados devem ser armazenados obrigatoriamente em um

datalake(Bucket ) , DW(BigQuery) ou em ambos.

2.8. Os Dataframe(s) resultante(s) deve(m) estar em uma coleção do

MongoDB atlas (informar a key de acesso ao cluster)

2.9. Deve ser feita no mínimo 4 análises dentro do Big Query utilizando a

linguagem padrão SQL com a descrição das consultas feitas.

2.10. Deve ser criado no Looker Studio e no PowerBI um dashboard para

exibição gráfica dos dados tratados trazendo insights importantes

2.11. E deve ser demonstrado em um workflow simples (gráfico) as etapas de

ETL com suas respectivas ferramentas.

2.12. Fazer a documentação do projeto;

### **Introdução**

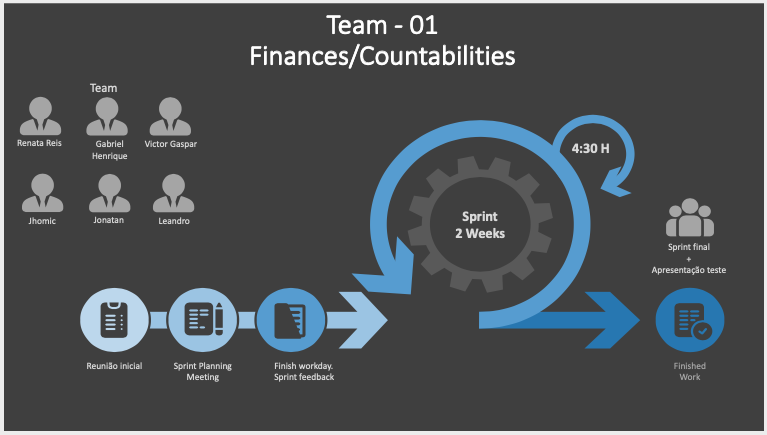
Nascida da audácia em explorar novos horizontes tecnológicos e desbravar fronteiras de inovação, a Rover Solutions é o resultado empolgante da união de seis analistas brasileiros altamente criativos. Combinando a elegância do roxo vibrante, o dinamismo do laranja intenso e a confiança do azul elétrico, a Rover Solutions emerge como uma constelação de talento, embarcando na jornada tecnológica para criar o futuro com originalidade e paixão.

Inspirados pelo slogan 'Desbravando Fronteiras Tecnológicas!', nossa equipe trabalha incansavelmente para criar soluções inovadoras que impactam positivamente o cenário tecnológico e além, enquanto permanecemos comprometidos em permanecer à frente da curva, como os verdadeiros pioneiros na exploração tecnológica.

É exatamente esse compromisso com a inovação e a busca incansável por novos horizontes que nos levou a aceitar o desafio de nos unirmos à SoulCode. Assim como nosso slogan sugere, estamos sempre desbravando fronteiras tecnológicas, trazendo nossa paixão pela originalidade e excelência para cada empreendimento que abraçamos. Ao embarcar nessa jornada conjunta com a SoulCode para a elaboração da mais excepcional análise de dados financeiros já vista entre as turmas de dados. Acreditamos que nossa colaboração resultará em uma abordagem inovadora e transformadora, refletindo a visão ousada que nos define.

## 

## Metodologia SCRUM



Fonte: Equipe Taurus

A equipe Rover seguiu todo o processo de elaboração do projeto utilizando da metodologia ágil SCRUM.

Horário de início das reuniões: 9:00h

Horário final das reuniões: 17:30h

Intervalos a cada 1h30 e uma pausa para o almoço às 12h.

Retorno do almoço: 13h30.

Sprint final do dia: 16:30

Cronograma

**9:00 - 9:40**

* Reunião da equipe
* Revisão dos objetivos do dia
* Discussão das dificuldades
* Distribuição das tarefas de cada membro do time

**9:41 - 11:50**

* Cada membro do time desenvolve o trabalho que foi direcionado individualmente e finaliza os itens da sua lista de pendências.

**11:51 – 12:00**

* Avisos do grupo antes da saída pro almoço.

**12:01**

* Almoço.

**13:30 – 13:35**

* Retorno do almoço.

**13:36**

* Os membros da equipe continuam trabalhando nos itens da lista de tarefas de sua responsabilidade.

**15:30 – 16:30**

* Comunicação de Intervalo.
* Os membros da equipe tomam um intervalo para esticar as pernas, comer um lanche e relaxar

**16:31 - 17:20**

* Trabalho em andamento
* Os membros da equipe concluem quaisquer itens restantes da lista de pendências

**17:21**

* Sprint de Fim do dia

**18:00**

* A equipe encerra o dia e se prepara para o dia seguinte.

## Datasets da Taurus, sua Origem

Figura 1 - Site Fundamentus

Fonte: https://www.fundamentus.com.br/Site Fundamentus, local da retirada de grandes dados da empresa Taurus


Figura 2 - Site Status Invest



Fonte: <https://statusinvest.com.br/acoes/tasa3>

1. Após escolhermos a Taurus como empresa para analisar utilizamos os sites do fundamentus e Status Invest para retirarmos dados originais para análise. A partir deles conseguimos 5 datasets, sendo 2 em formato pdf, 2 em csv e 1 em .xlsv.
2. Após analisarmos superficialmente o que esses datasets diziam, decidimos seguir com a análise de possíveis KPI’s.

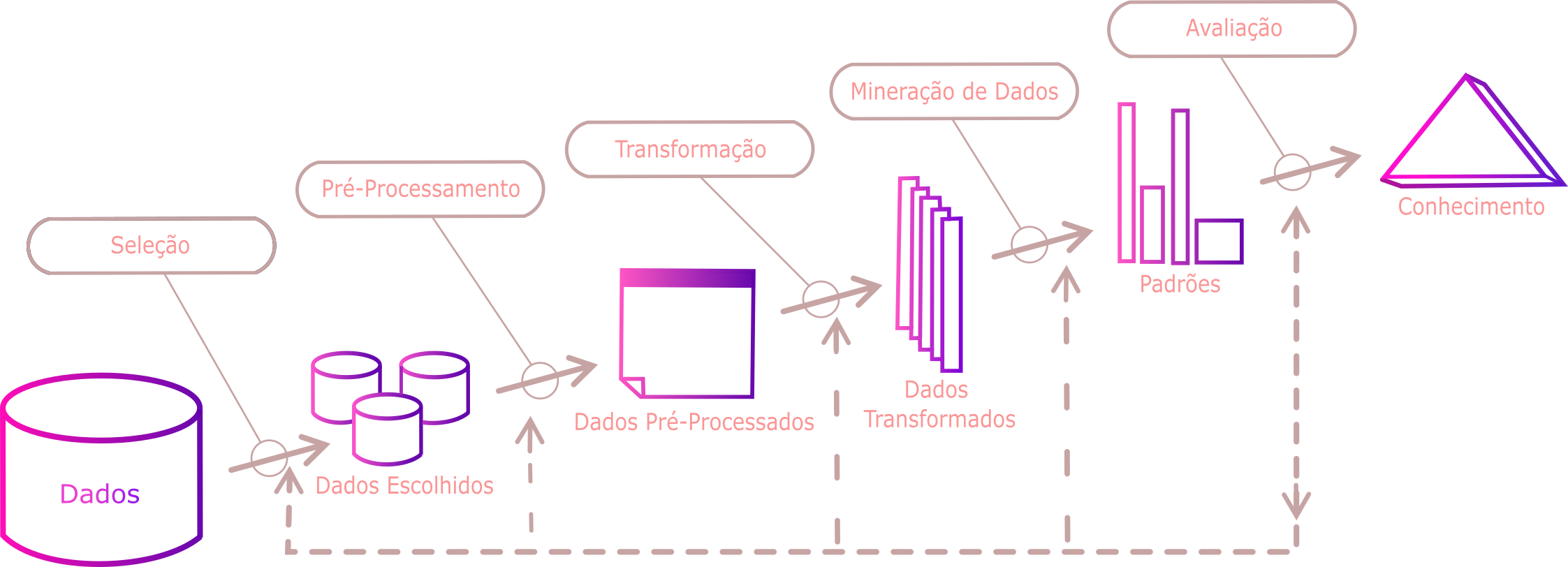
## KPI’s Definidos

* Balanço Patrimonial
* Ativo total da empresa;
* Endividamento/dívida líquida;
* Patrimônio líquido;
* Reserva de lucros;
* Demonstrativo de Resultados
* Resultado bruto;
* Receita líquida;
* margens de contribuição;
* Resultado financeiro /lucro líquido;
* Fluxo de Caixa
* EBITDA
* Depreciação
* Capital de giro

A partir da análise que a equipe Rover efetuou foi identificado a necessidade de buscarmos as 3 principais Kpi’s (Balanço patrimonial, Dre e Fluxo de caixa) e suas subcategorias para termos insights efetivos na análise da premissa.

## Metodologia KDD

Figura 3 - Método KDD

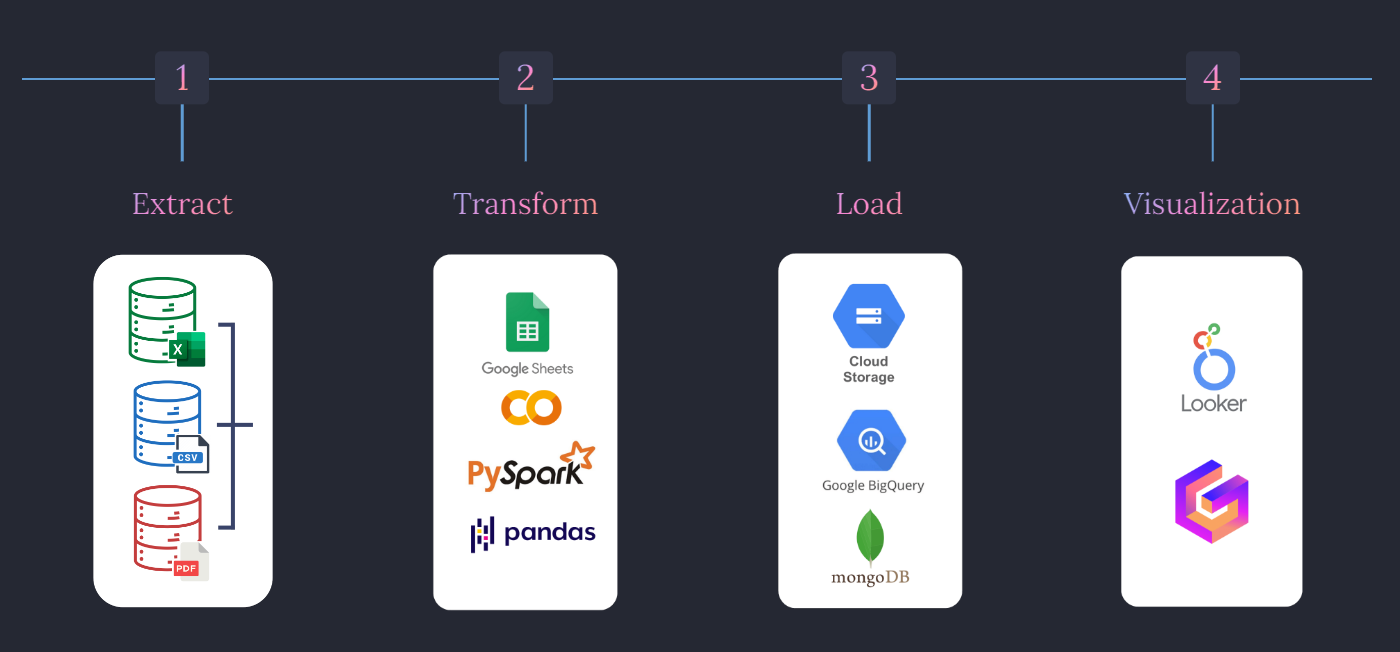


Fonte: Equipe Taurus

Nossa equipe escolheu a metodologia KDD porque oferece uma abordagem estruturada e eficaz para explorar nossos dados. A KDD, com seu grande volume de informações, nos ajuda a encontrar padrões e insights ocultos, o que nos permite tomar decisões mais poderosas e encontrar oportunidades de melhoria. Essa abordagem sistemática aumenta nossa capacidade de inovação e tomada de decisões estratégicas, envolvendo dados em conhecimento acionável.

## Workflow

Figura 4 - Workflow Rover



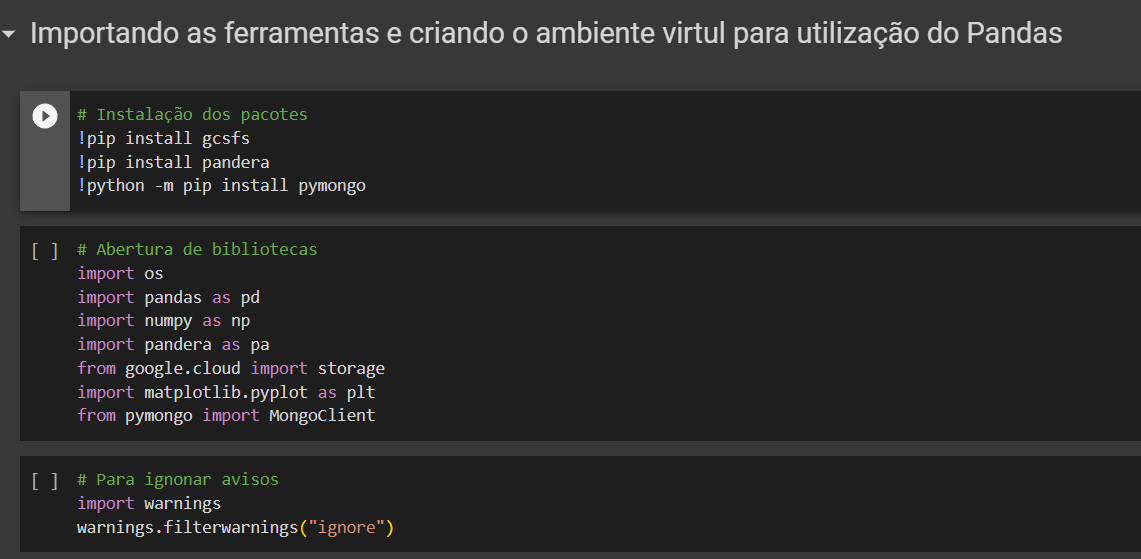
Fonte: Equipe Taurus

Nosso processo de trabalho é organizado em um processo seguro. O processo começa com uma etapa de garantia, na qual coletamos dados pertinentes de várias fontes. Em seguida, os dados são organizados, limpos e preparados para análise durante uma fase de transformação. Após a transformação, os dados são carregados em nosso sistema e estão prontos para uso. A visualização dos dados é a última etapa. Isso permite que os insights e padrões emergentes sejam apresentados de forma fácil de entender e forneçam uma base sólida para tomada de decisões controladas.

## Análise de dados: Etapa Pandas

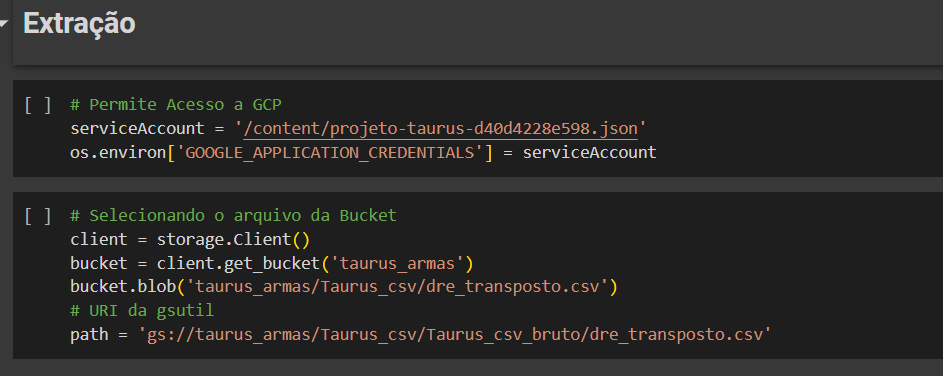
### **Dataframe DRE:**

Figura 5 - Instalação dos pacotes no Colab



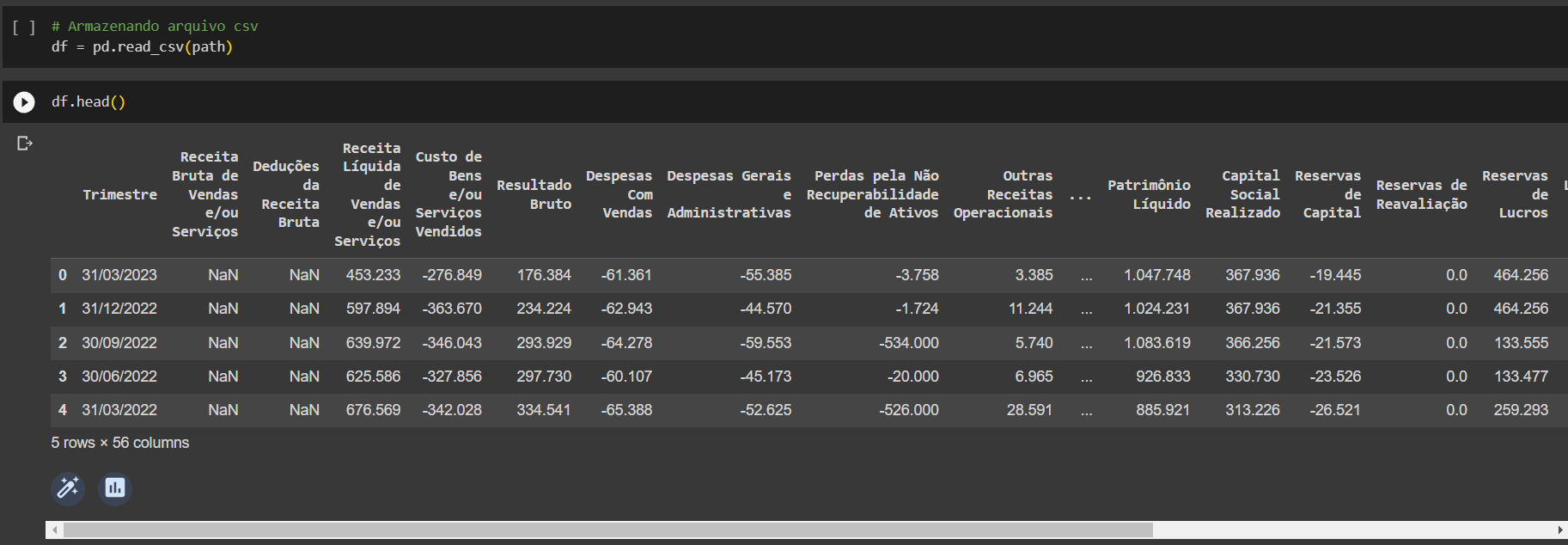
Fonte: Equipe Taurus

Figura 6 - Extração dos dados salvos na Bucket



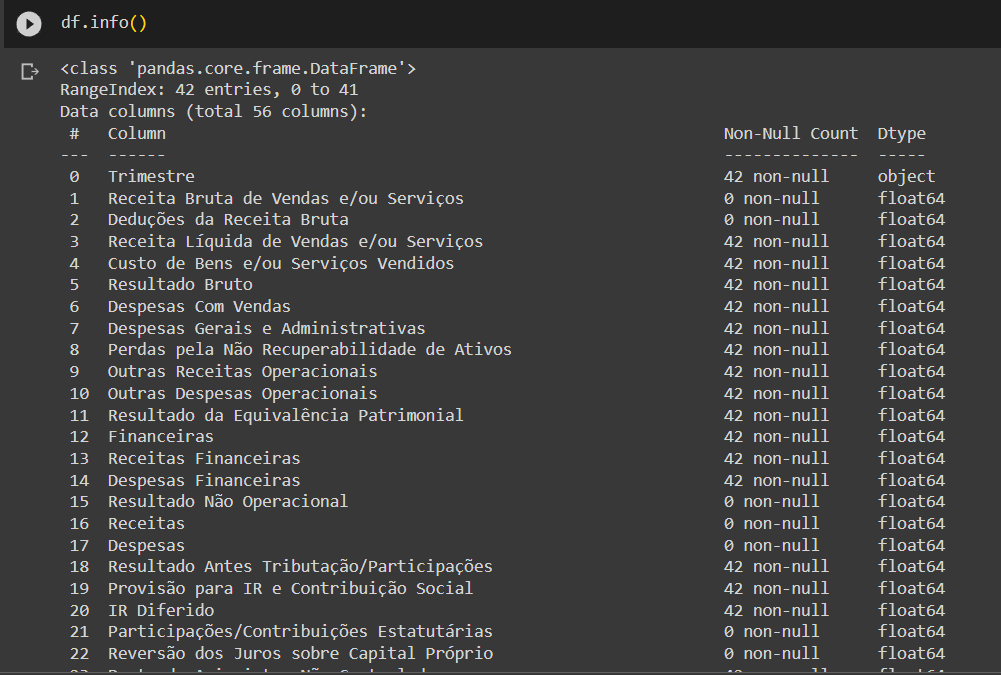
Fonte: Equipe Taurus

Figura 7 - Analisando o cabeçalho do dataframe DRE



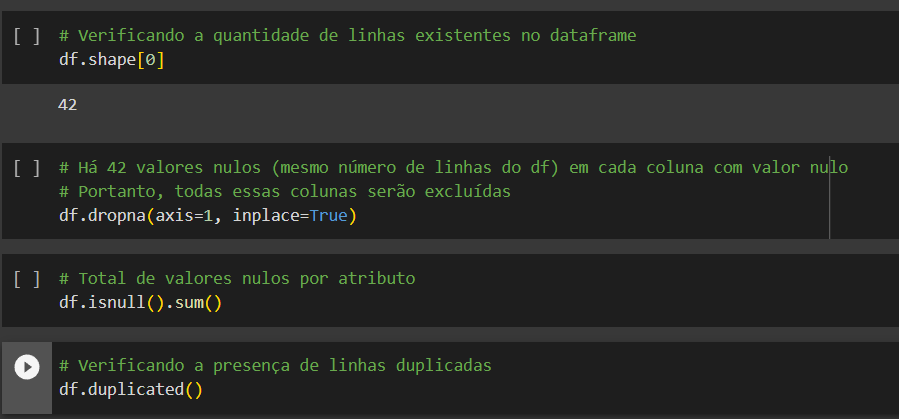
Fonte: Equipe Taurus

Figura 8 - Verificando as informações dos atributos do dataframe DRE



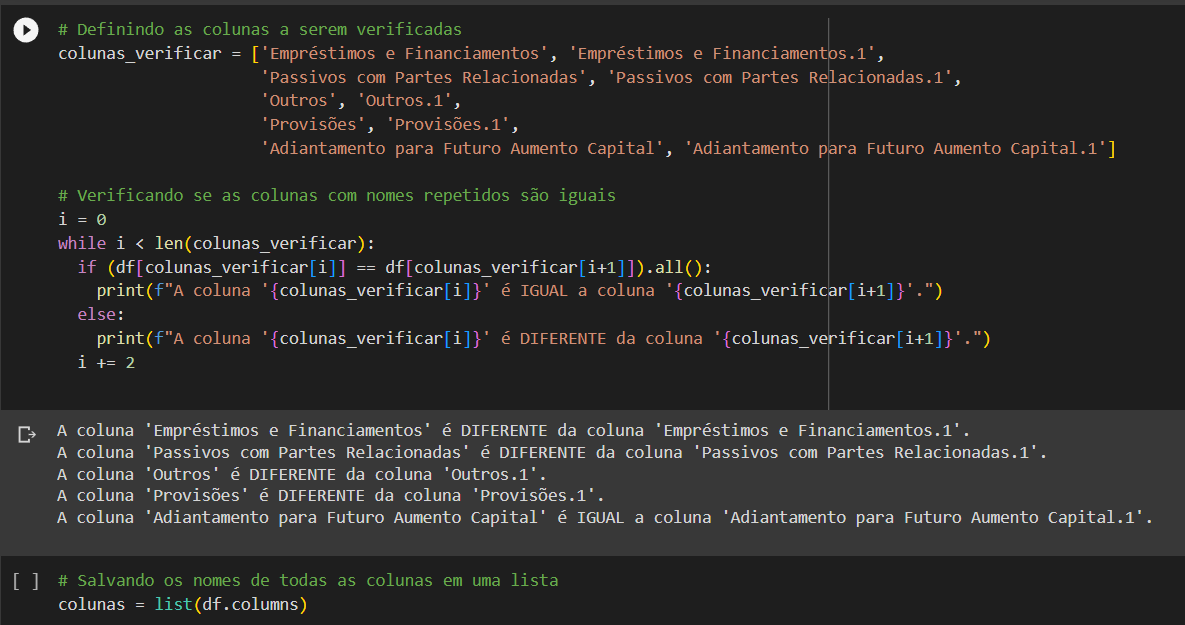
Fonte: Equipe Taurus

Figura 9 - Analisando existência de nulos e linhas duplicadas



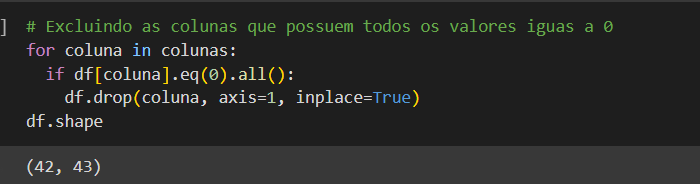
Fonte: Equipe Taurus

Figura 10 - Definindo se as colunas são iguais



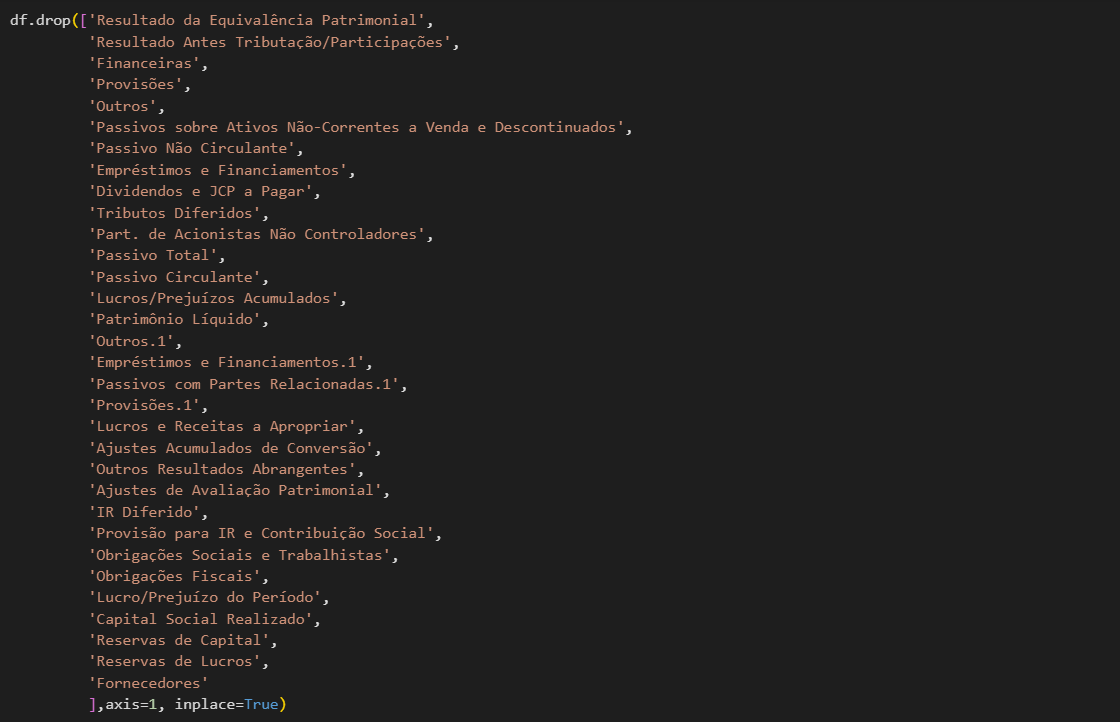
Fonte: Equipe Taurus

Figura 11 - Excluindo colunas com valores (0)



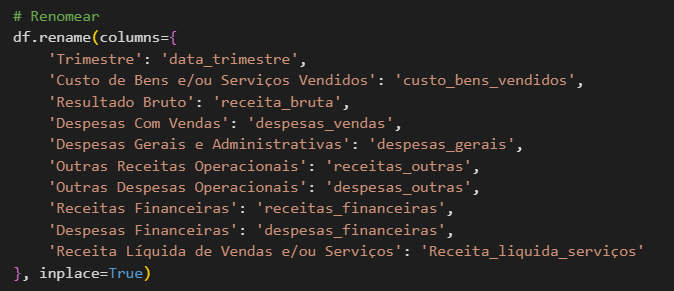
Fonte: Equipe Taurus

Figura 12 - Dropando as colunas que não irão ser utilizadas na analise



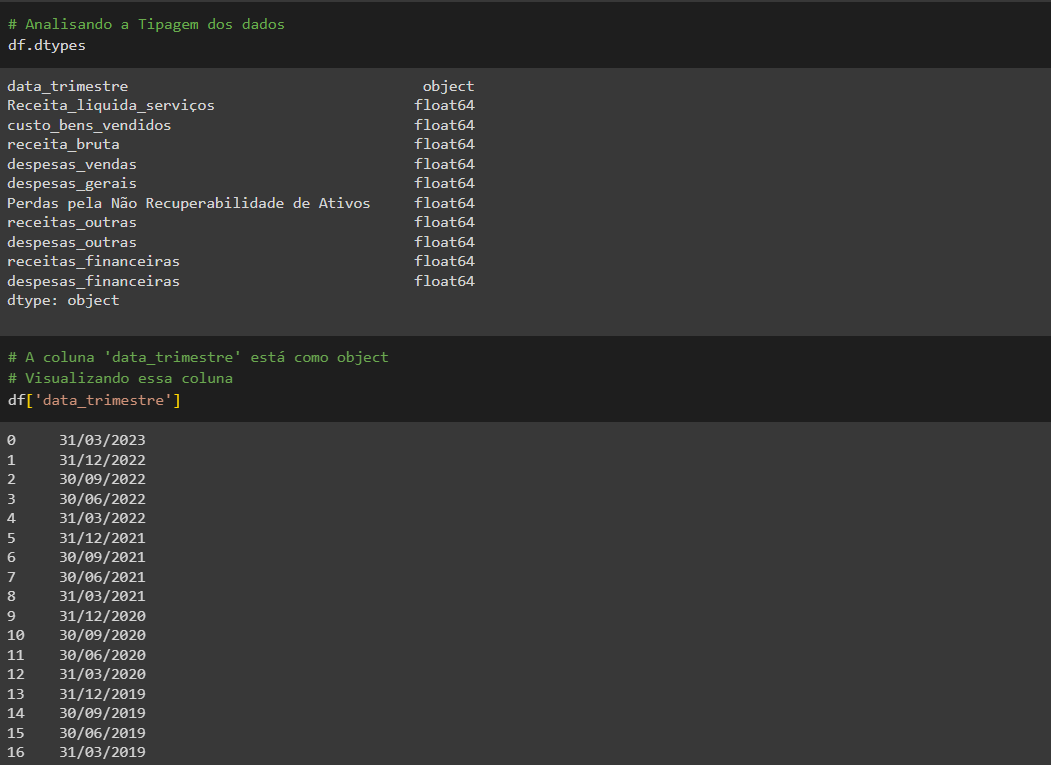
Fonte: Equipe Taurus

Figura 13 - Renomeando



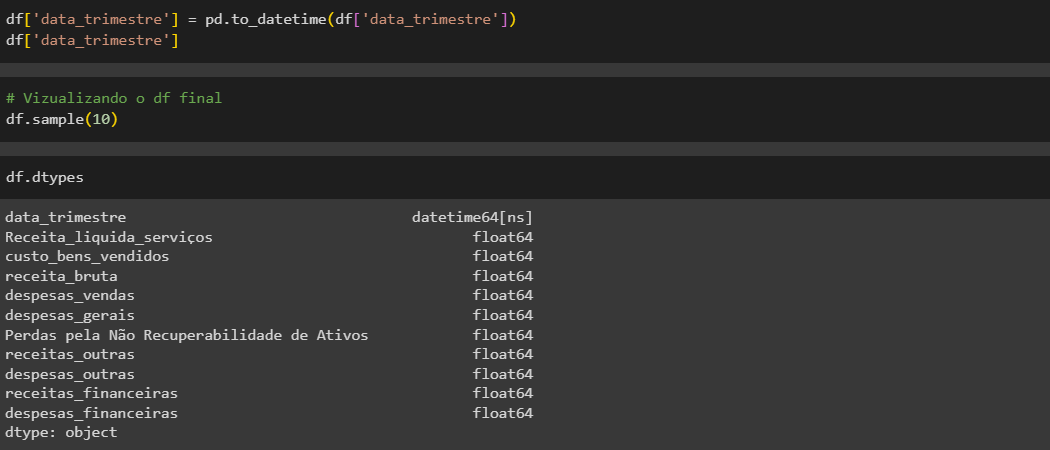
Fonte: Equipe Taurus

Figura 14 - Tipagem e análise das colunas com type á ser alterado



Fonte: Equipe Taurus

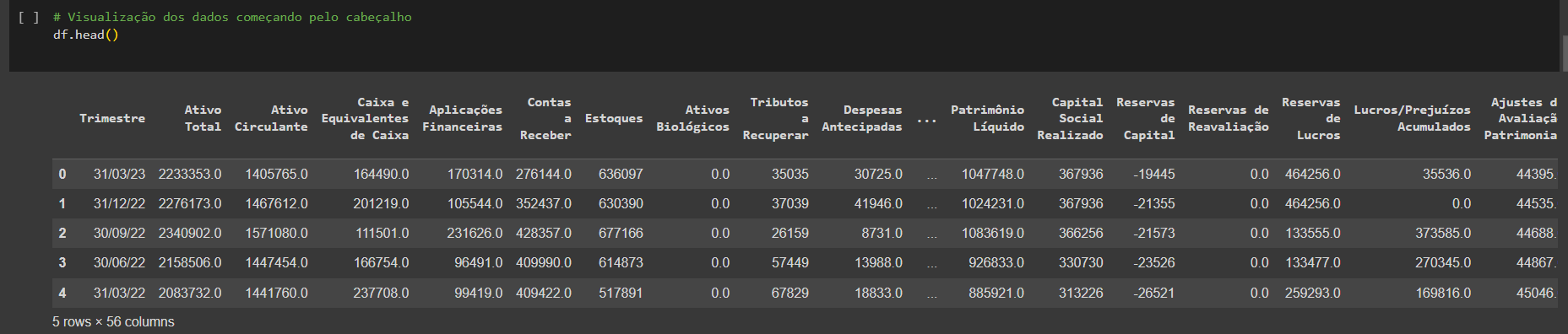
Figura 15 - Convertendo o tipo de dados e visualizando essa alteração



Fonte: Equipe Taurus

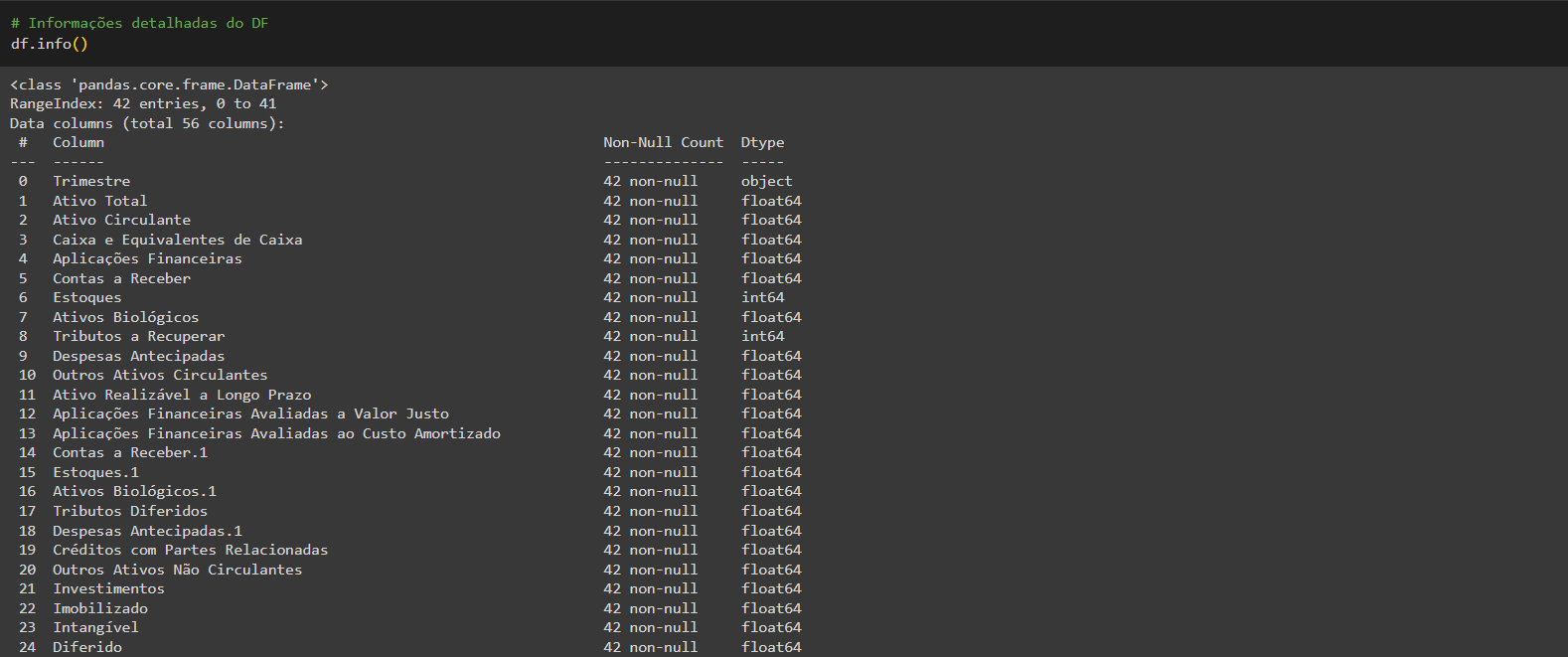
### **Dataframe Balanço Patrimonial:**

Figura 16 - Analisando o cabeçalho do dataframe Balanço Patrimonial



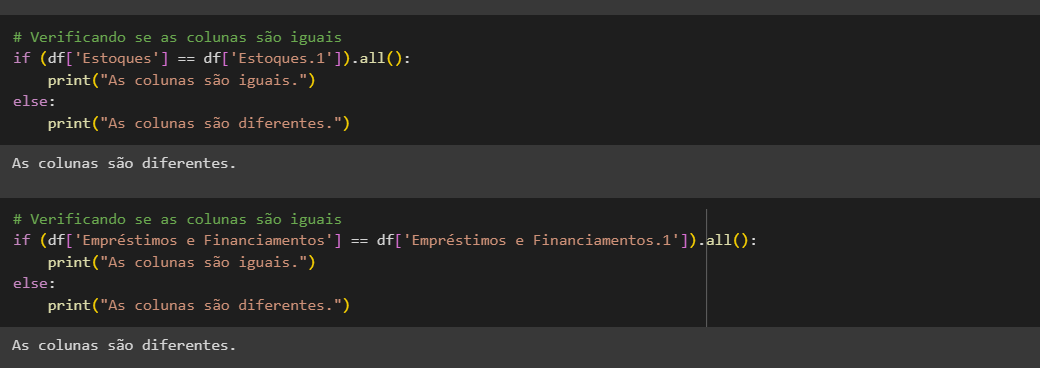
Fonte: Equipe Taurus

Figura 17 - Verificando os dados de cada atributo



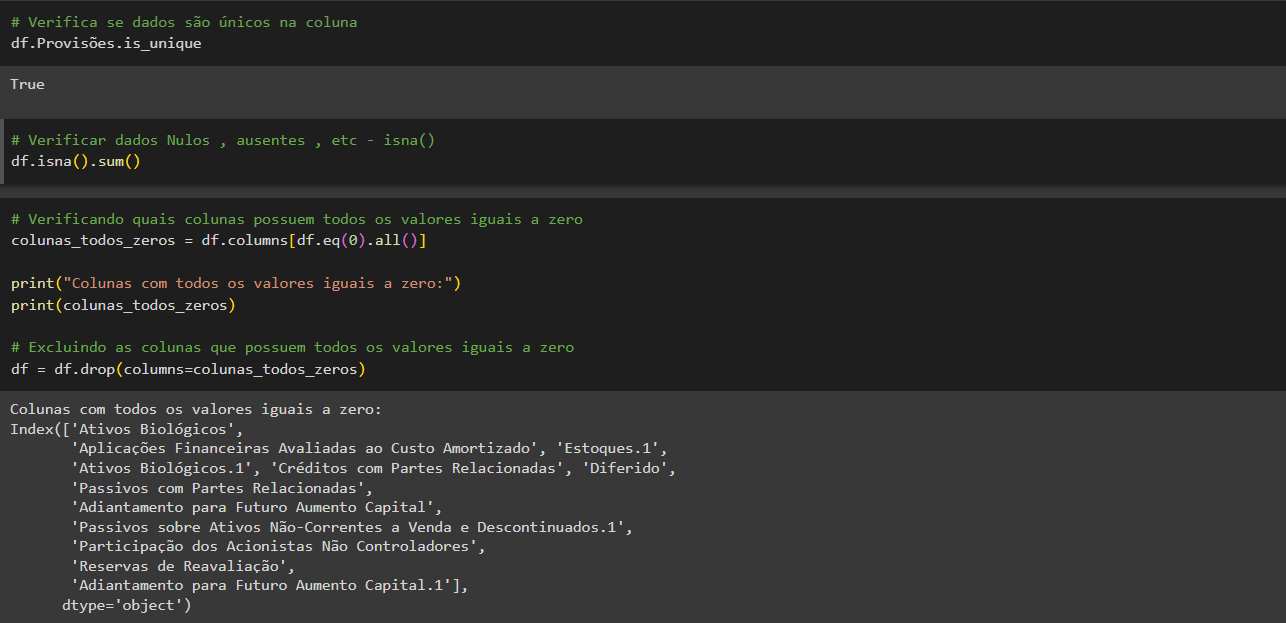
Fonte: Equipe Taurus

Figura 18 - Verificando igualdade de dados entre colunas



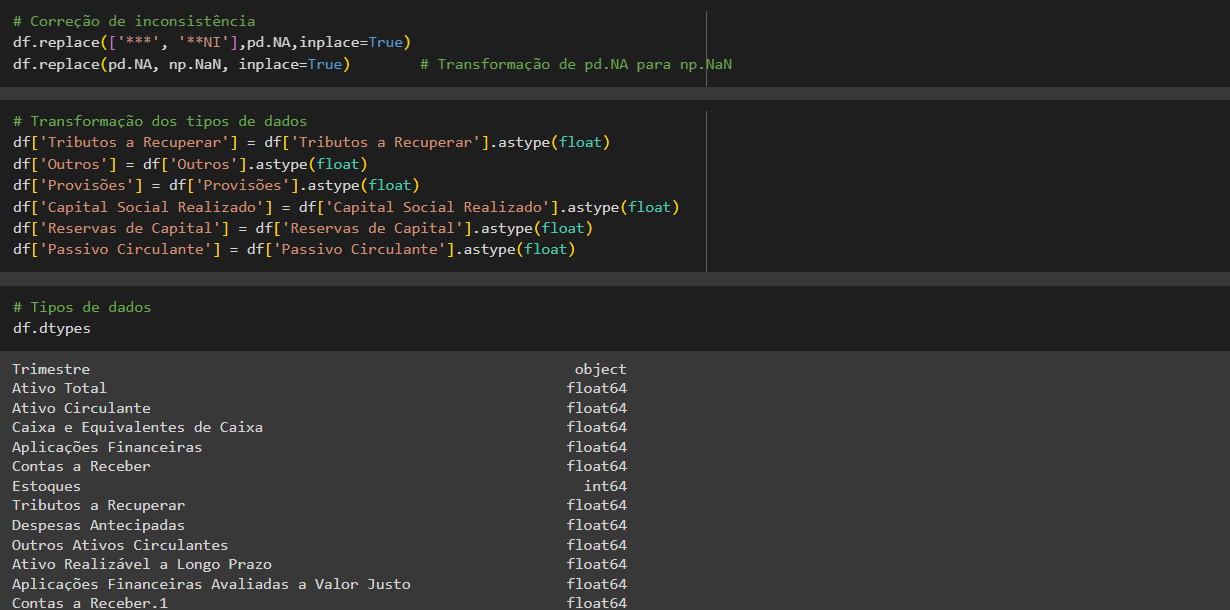
Fonte: Equipe Taurus

Figura 19 - Verificando valores únicos, nulos, ausentes e excluindo os que contém valor 0



Fonte: Equipe Taurus

Figura 20 - Correção de inconsistências e transformação dos tipos de dados



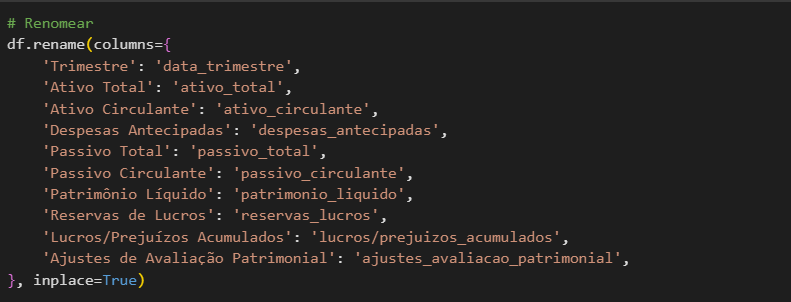
Fonte: Equipe Taurus

Figura 21 - Drop das colunas não necessárias para a análise



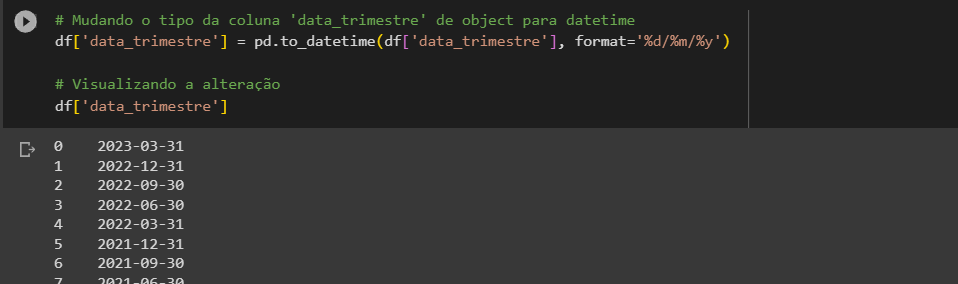
Fonte: Equipe Taurus

Figura 22 - Renomeação das colunas



Fonte: Equipe Taurus

Figura 23 - Alteração da coluna data\_trimestre do tipo object para datetime

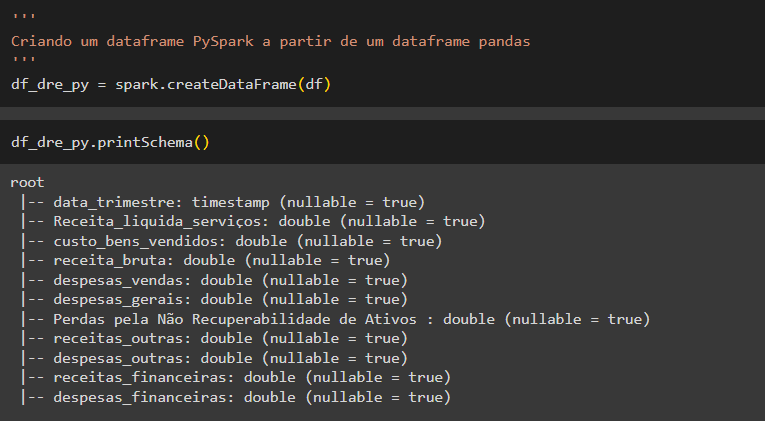


Fonte: Equipe Taurus

## Análise de dados: Etapa Pyspark

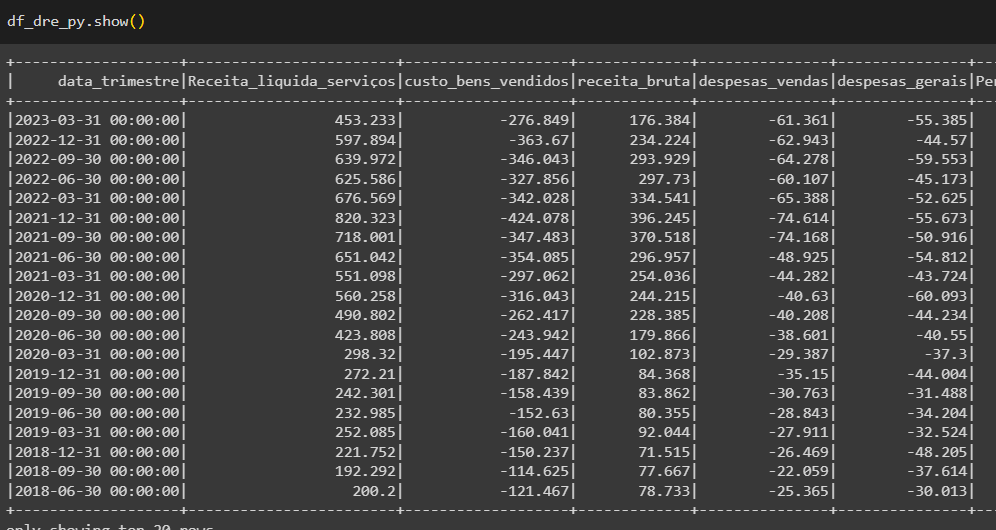
### **Dataframe DRE:**

Figura 24 - Convertendo o df pandas para pyspark



Fonte: Equipe Taurus

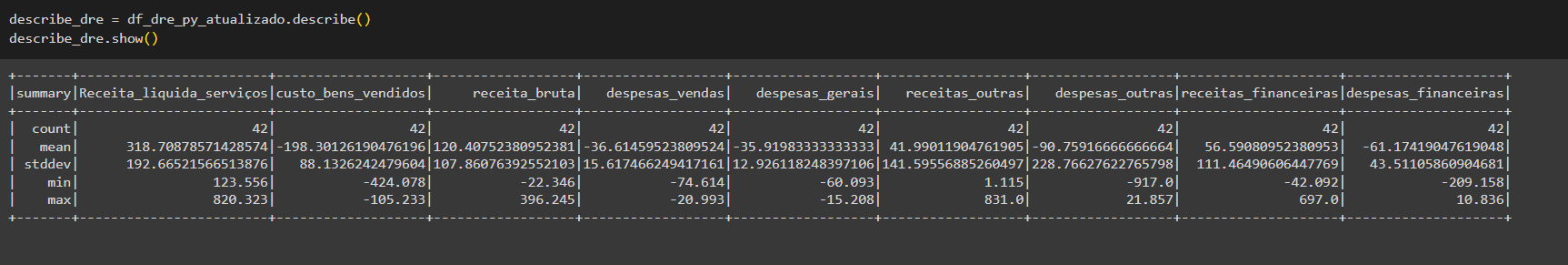
Figura 25 - Analisando o Dataframe Pyspark DRE



Fonte: Equipe Taurus

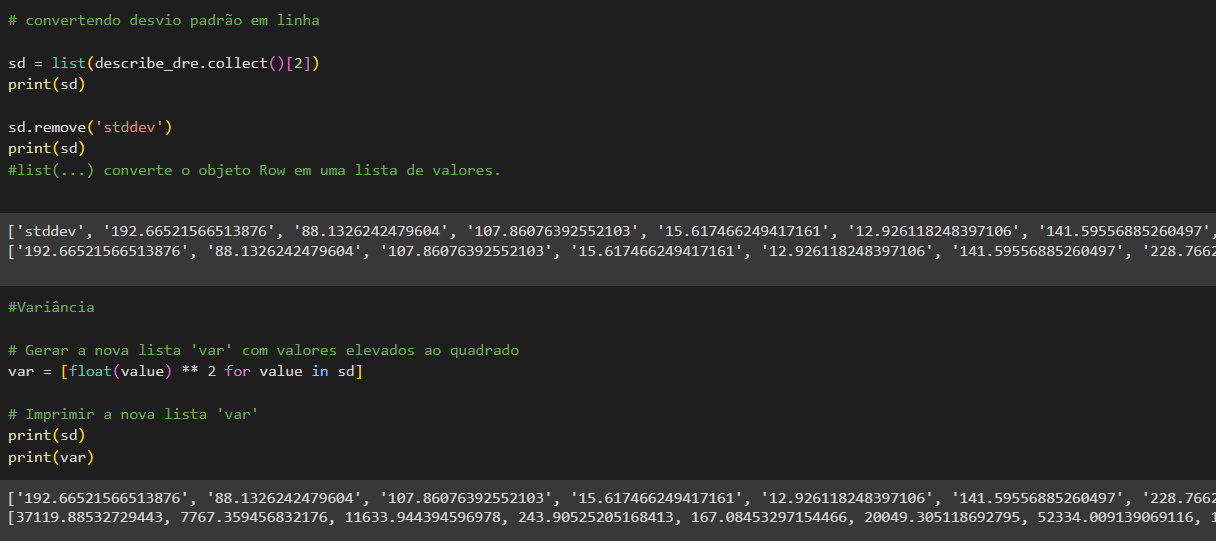
#### Análise descritiva DRE

Figura 26 - Criando o describe do df DRE



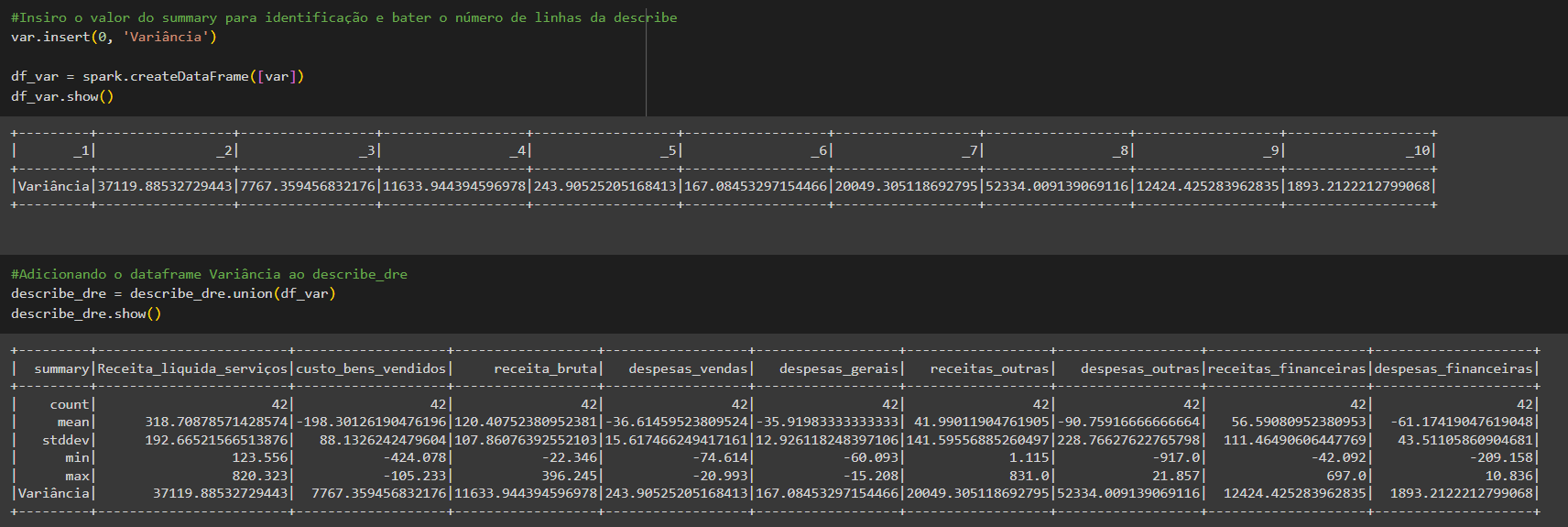
Fonte: Equipe Taurus

Figura 27 - Retirando a linha stddev para cálculo da variância



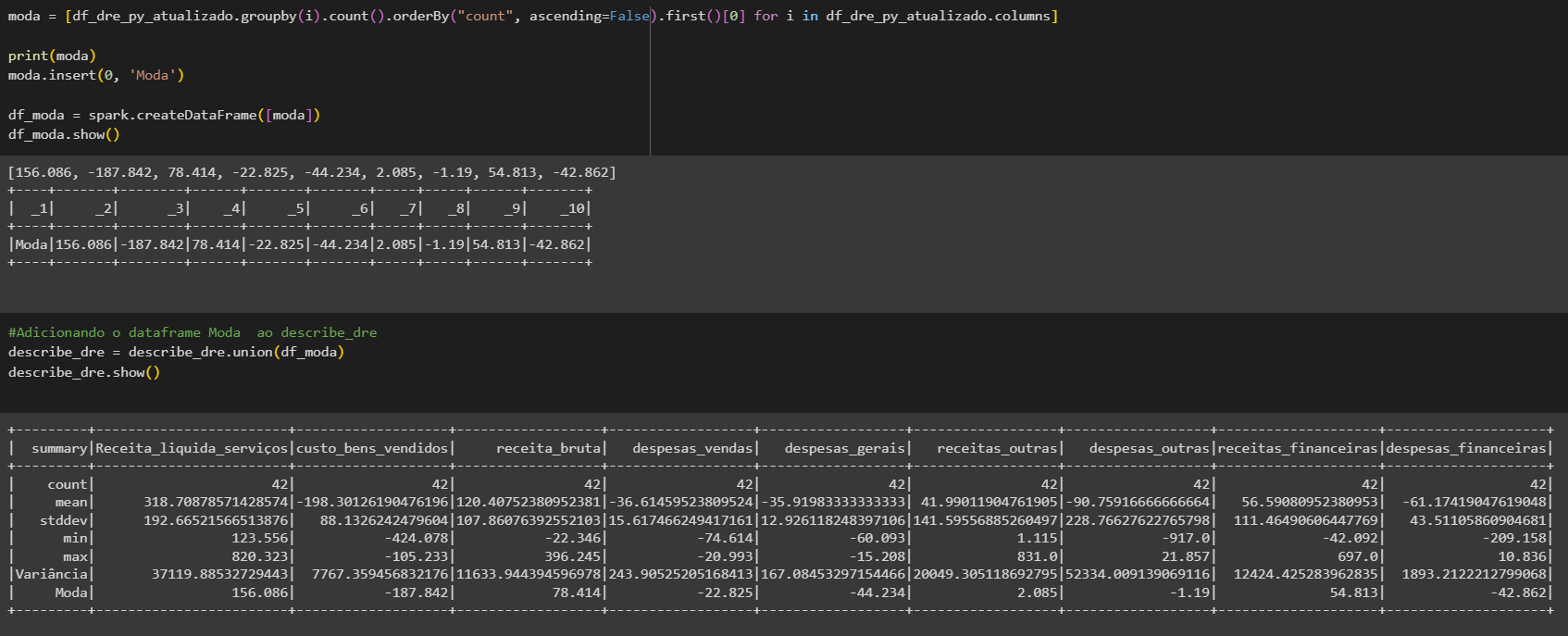
Fonte: Equipe Taurus

Figura 29 - Criação do df var e junção ao df describe\_dre



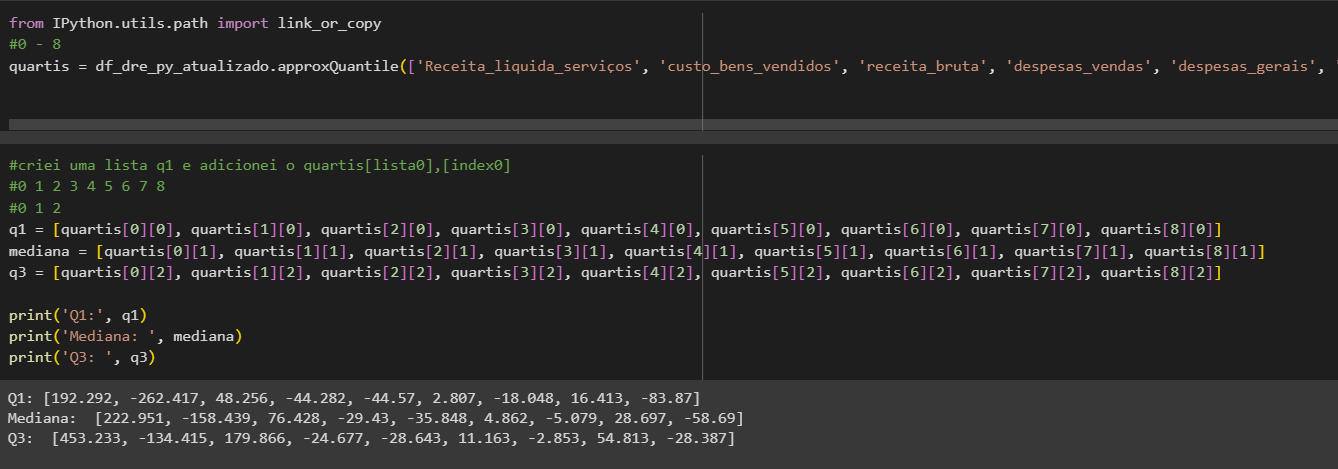
Fonte: Equipe Taurus

Figura 30 - Calculando moda, criando sua df e acrescentando ao df describe\_dre



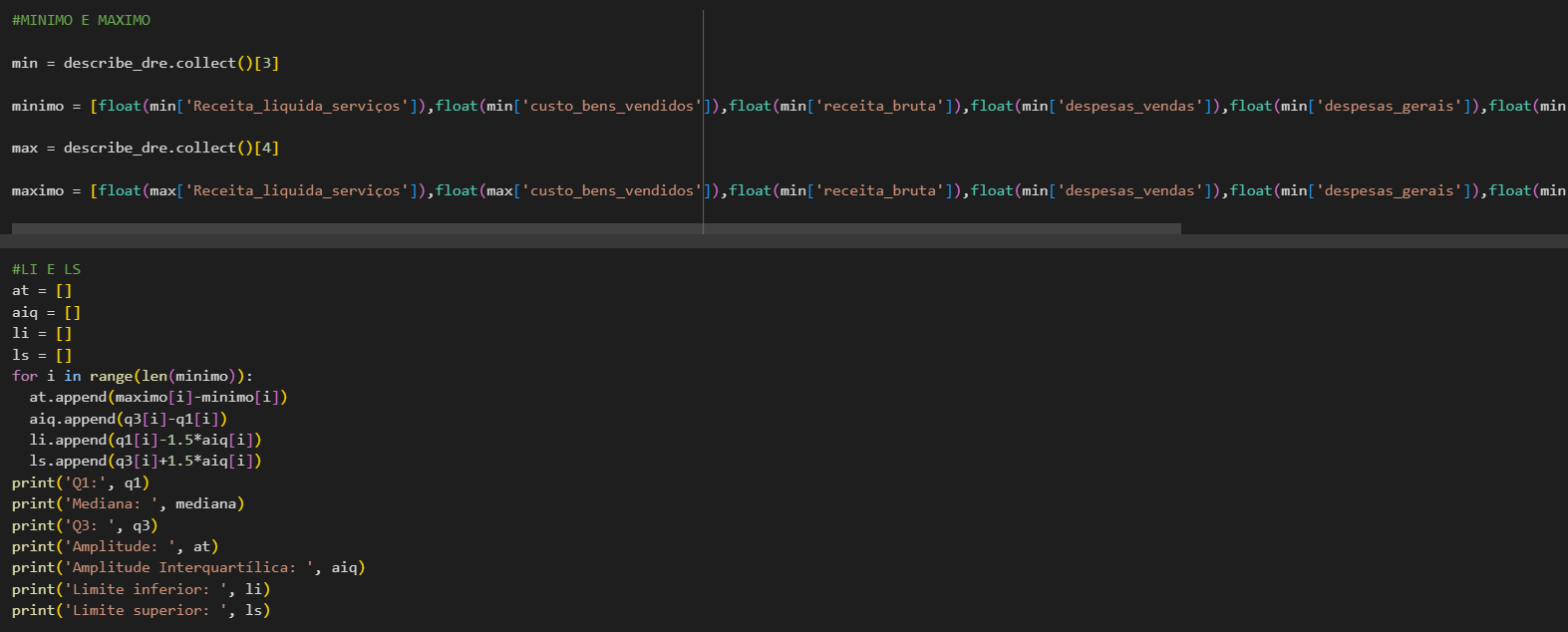
Fonte: Equipe Taurus

Figura 31 - Cálculo dos Quartis da df



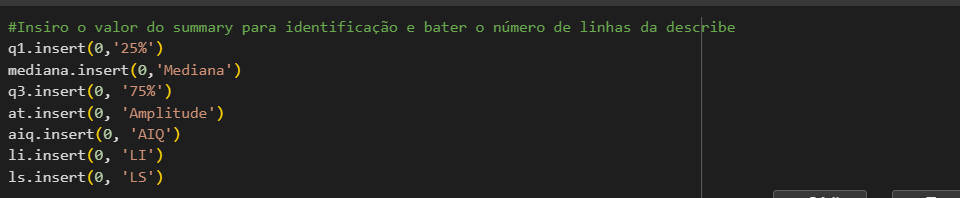
Fonte: Equipe Taurus

Figura 32 - Tratamento dos valores mínimo, máximo, AT, AIQ, Li e Ls



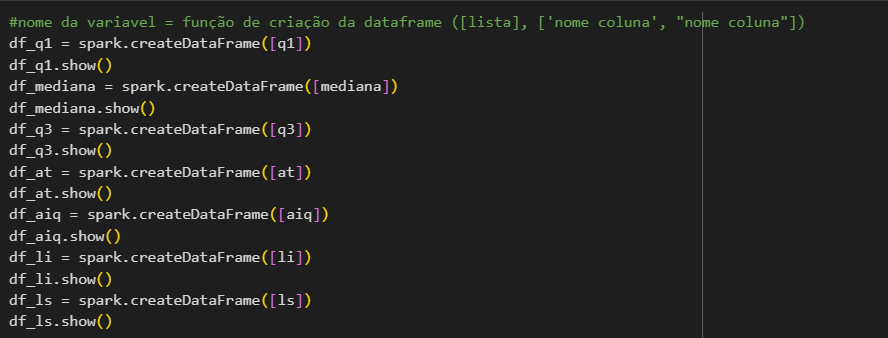
Fonte: Equipe Taurus

Figura 33 - Adicionando ao index 0 valores string/object



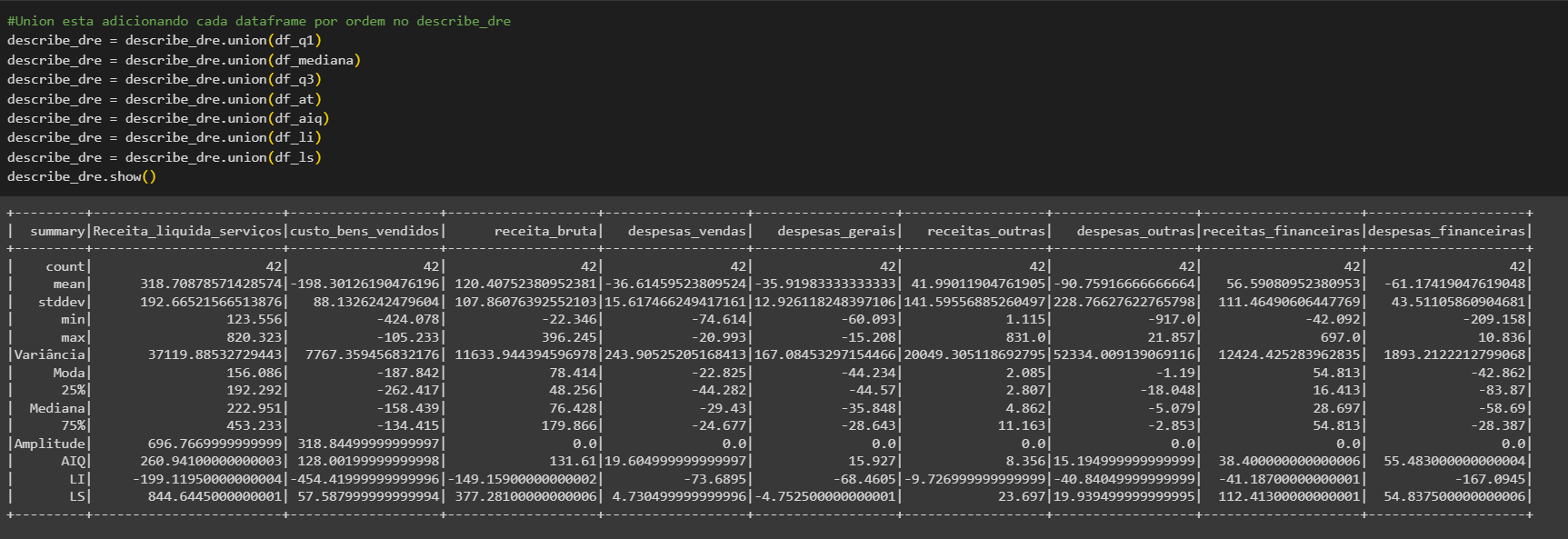
Fonte: Equipe Taurus

Figura 34 - Criando um df para cada dado descritivo



Fonte: Equipe Taurus

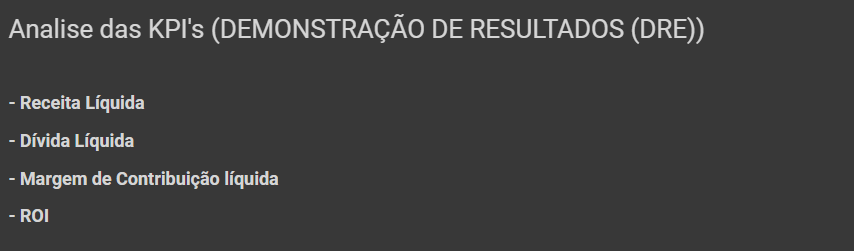
Figura 35 - União de todos os df resultantes e df describe\_dre final



Fonte: Equipe Taurus

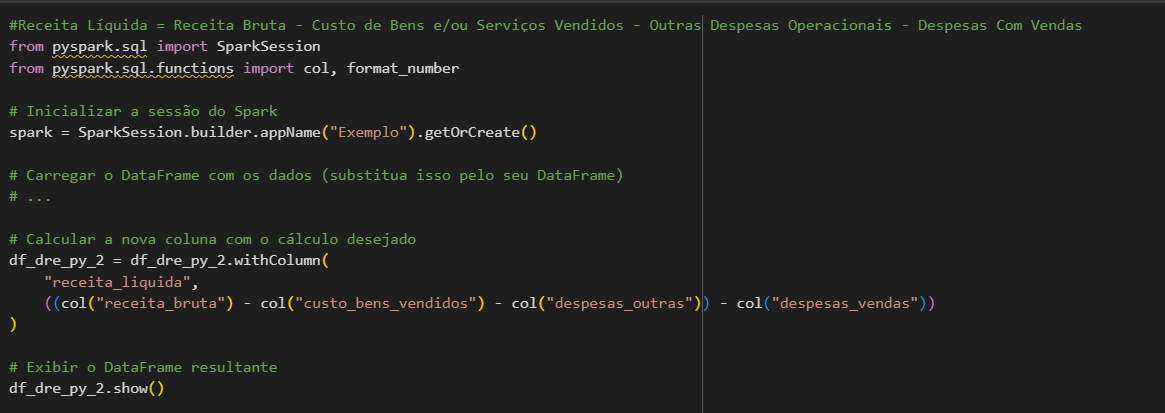
#### Análise KPI’s DRE

Figura 36 - Análise das kpi’s



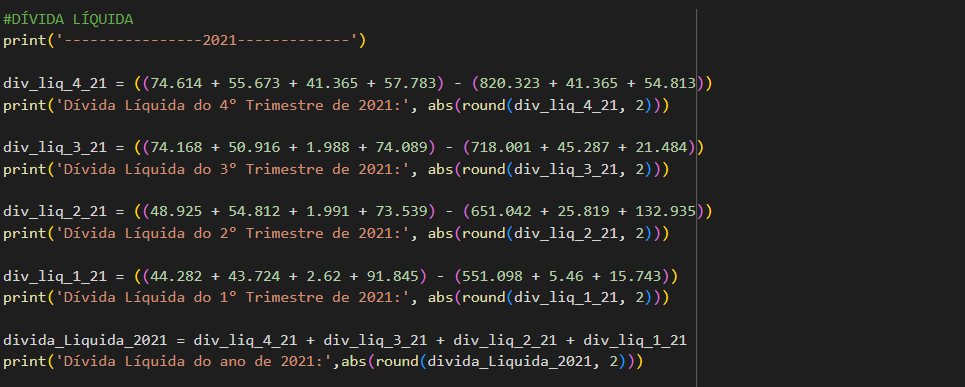
Fonte: Equipe Taurus

Figura 37 - Criando uma nova coluna “Receita Líquida” no df



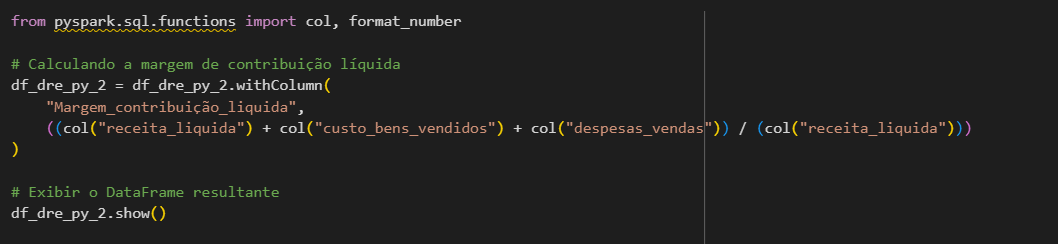
Fonte: Equipe Taurus

Figura 38 - Calculando a Dívida Líquida por trimestre



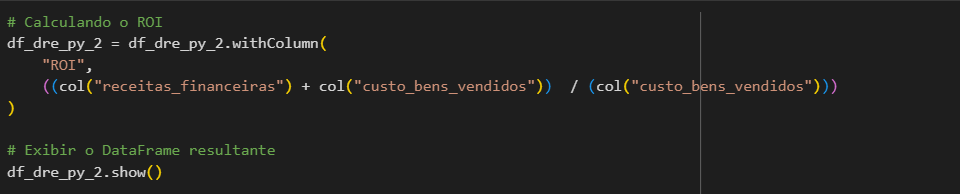
Fonte: Equipe Taurus

Figura 39 - Criando nova coluna “Margem\_contribuicao\_liquida”



Fonte: Equipe Taurus

Figura 40 - Criando nova coluna “ROI”



Fonte: Equipe Taurus

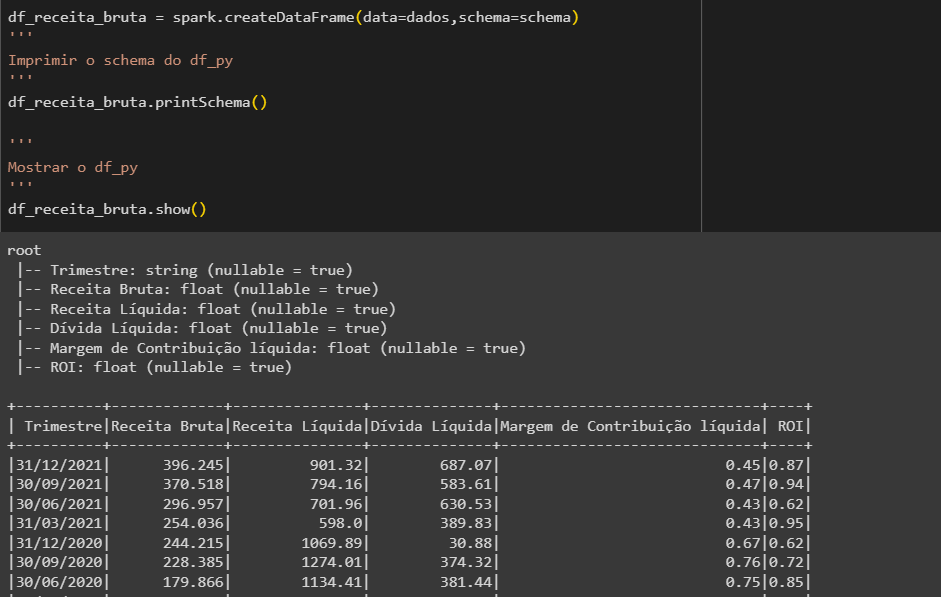
#### Análise Comparativa de receitas brutas dos trimestres de 2019,2020 e 2021

Figura 41 - Criação de um df pyspark com metodo schema

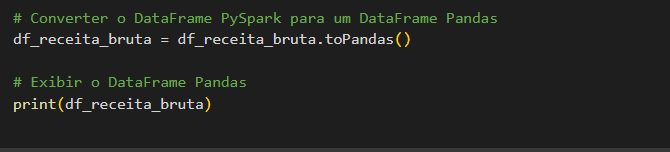


Fonte: Equipe Taurus

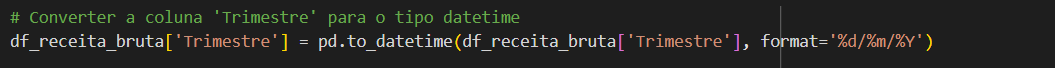
Figura 41.1 - Impressão e visualização do df\_receita\_bruta



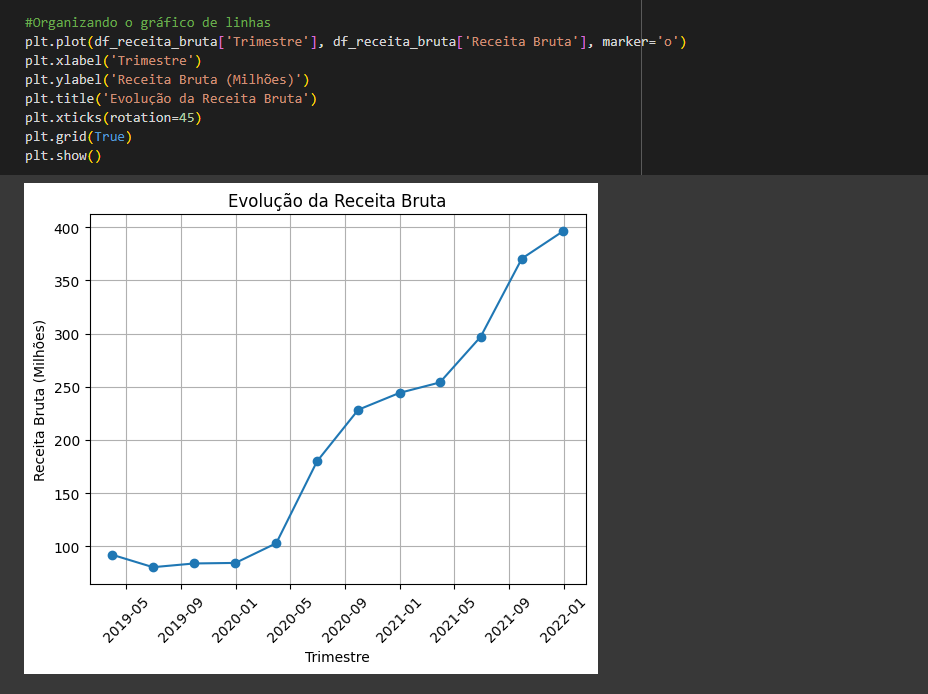
Fonte: Equipe Taurus

Figura 42 - Convertendo df\_receita\_bruta pyspark para pandas

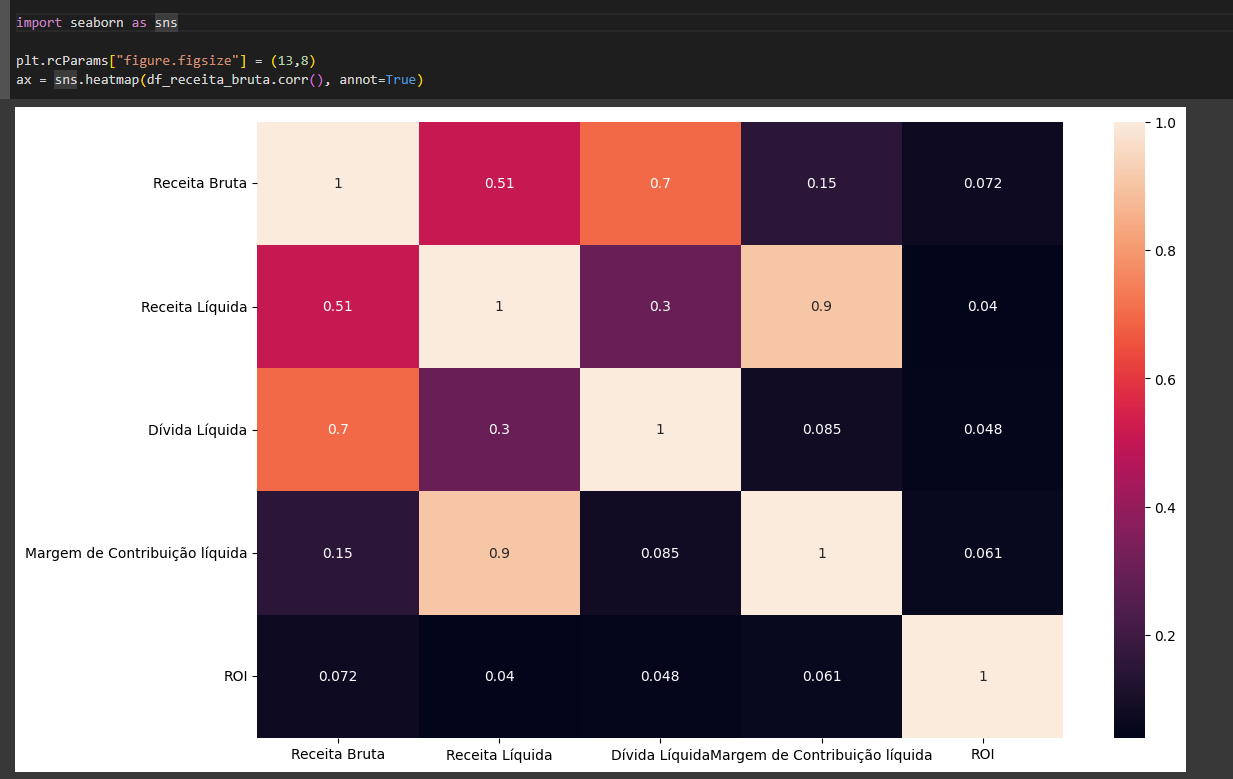
Fonte: Equipe Taurus

Figura 43 - Convertendo o tipo da coluna trimestre para datatime 

Fonte: Equipe Taurus

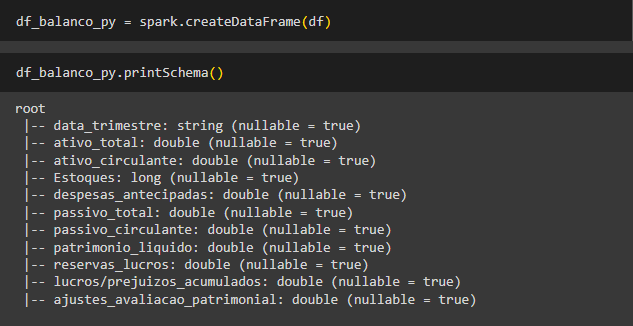
Figura 44 - Criação do gráfico comparativo da receita bruta por trimestre

Fonte: Equipe Taurus

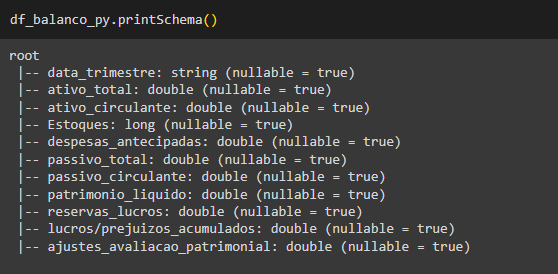
Figura 45 - Gráfico de correlação

Fonte: Equipe Taurus

### **Dataframe Balanço Patrimonial:**

Figura 46 - Convertendo df balanço patrimonial pandas para pyspark

Fonte: Equipe Taurus

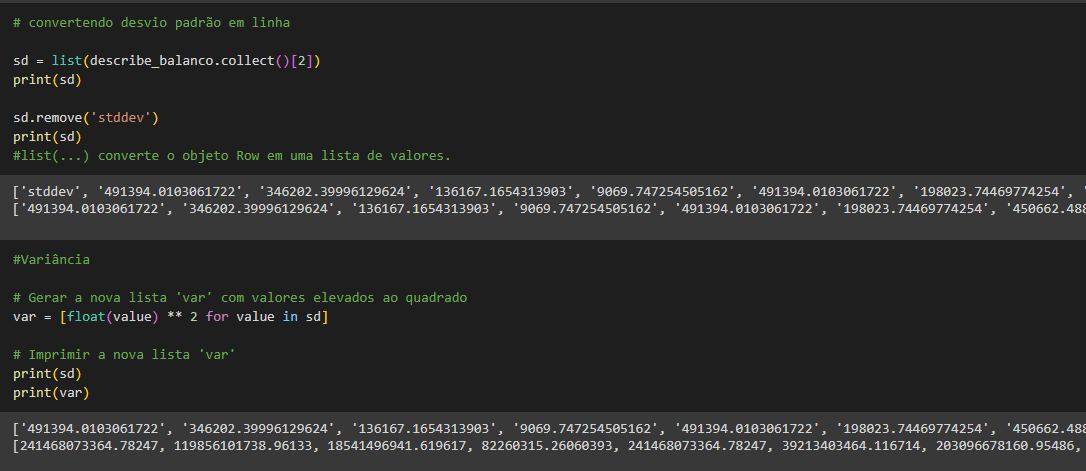
Figura 47 - Visualizando o schema do df\_balanco\_py

Fonte: Equipe Taurus

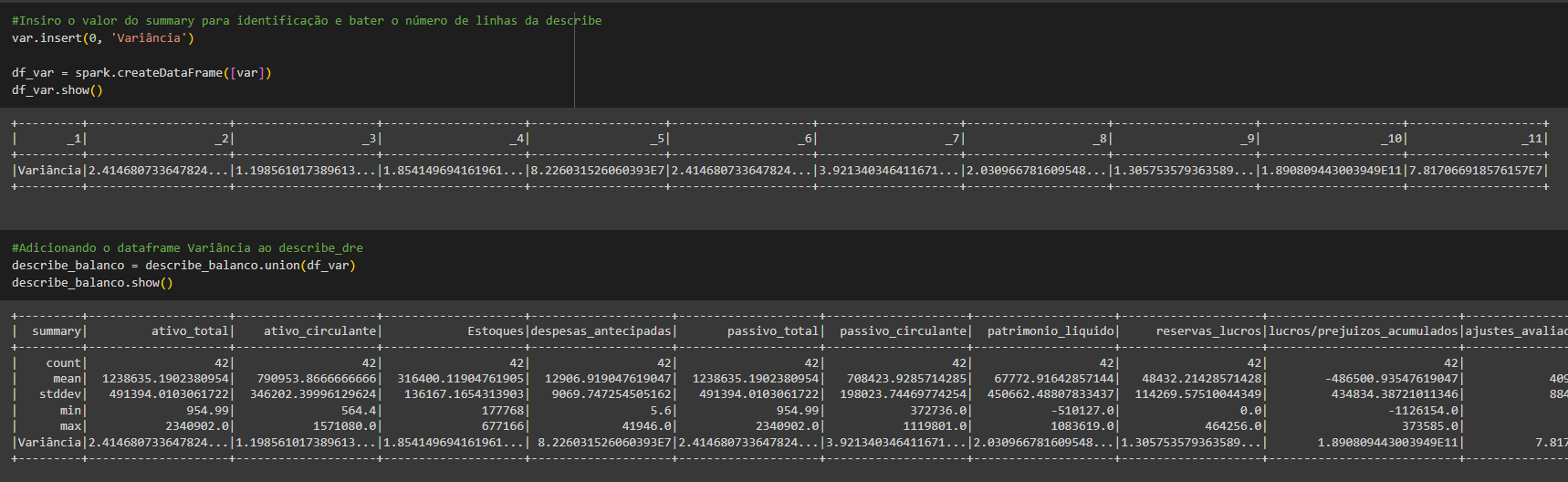
#### Análise descritiva DRE

Figura 48 - criando um describe do df\_balanco\_py

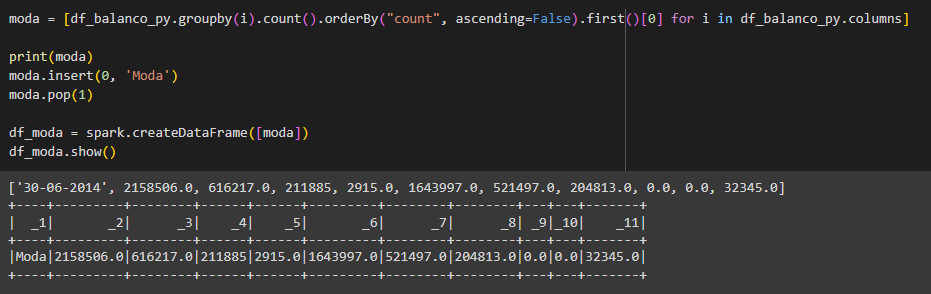
Fonte: Equipe Taurus

Figura 49 - Retirando a linha stddev e calculando a variância apartir dela

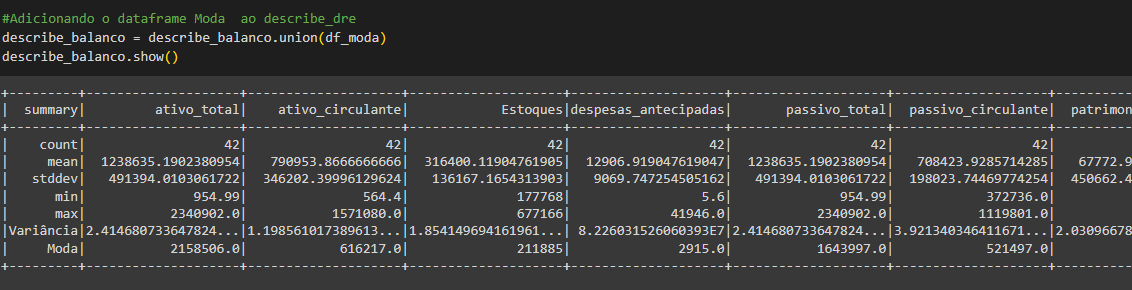
Fonte: Equipe Taurus

Figura 50 - Criando um df var e adicionando ao df describe

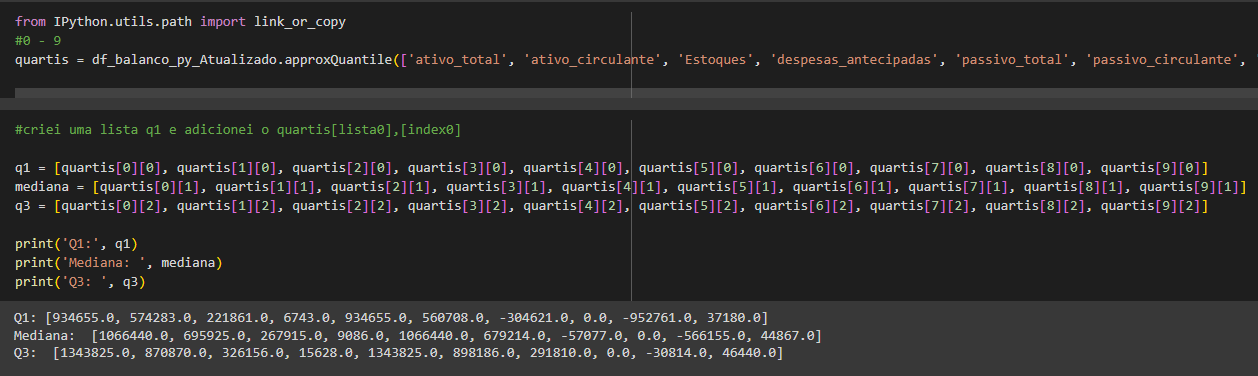
Fonte: Equipe Taurus

Figura 51 - Calculando a moda do df\_balanco\_py e criando sua df

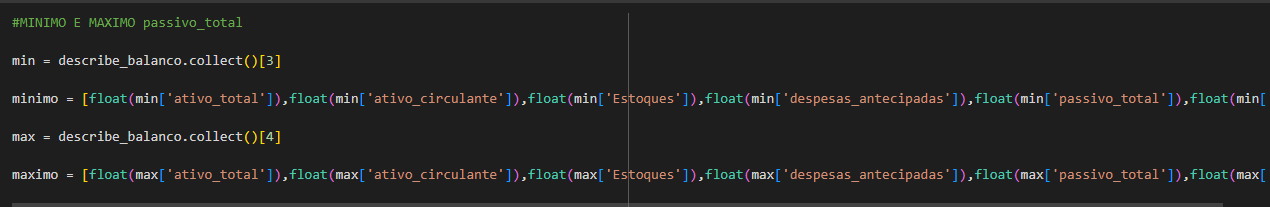
Fonte: Equipe Taurus

Figura 52 - df describe unificado com a moda

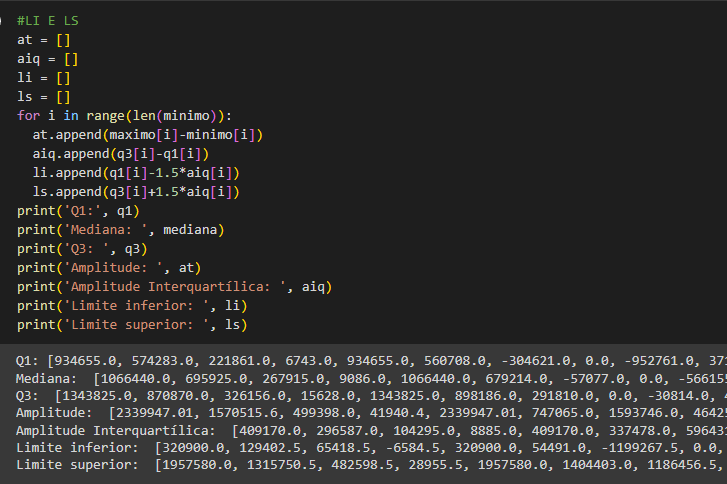
Fonte: Equipe Taurus

Figure 53 - Calculando os quartis da df balanco

Fonte: Equipe Taurus

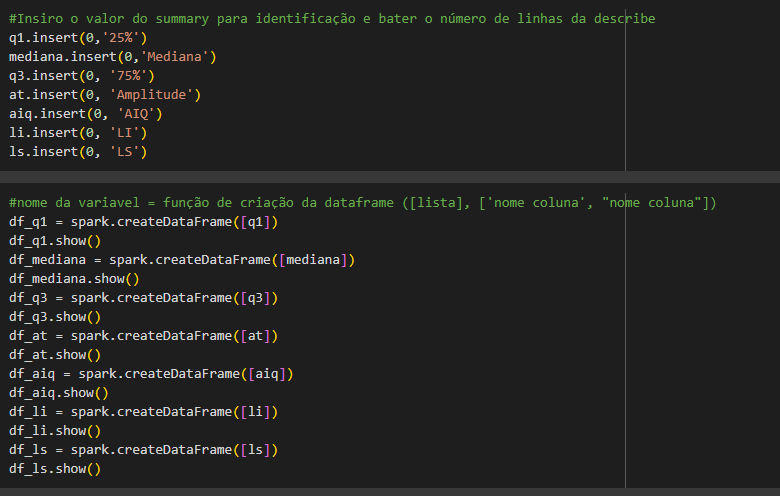
Figura 54 - Tratando os valores mínimo e maximo

Fonte: Equipe Taurus

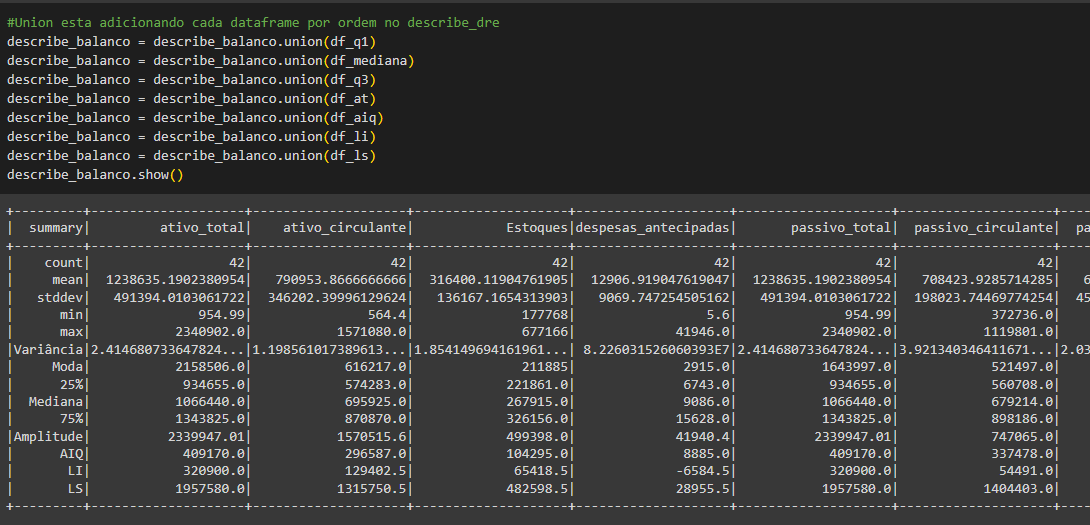
Figura 55 - Calculando o AT, AIQ, Li e Ls

Fonte: Equipe Taurus

Figura 56 - Inserindo uma string/object no index 0 e criando suas respectivas df



Fonte: Equipe Taurus

Figura 57 - df describe\_balanco com todos os dados unificados 

Fonte: Equipe Taurus

#### Análise KPI’s Balanço Patrimonial

Figura 58 - Kpi’s

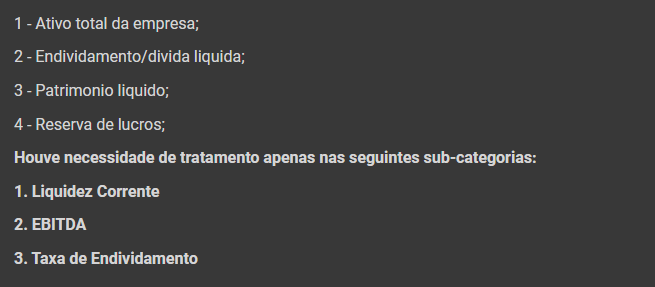
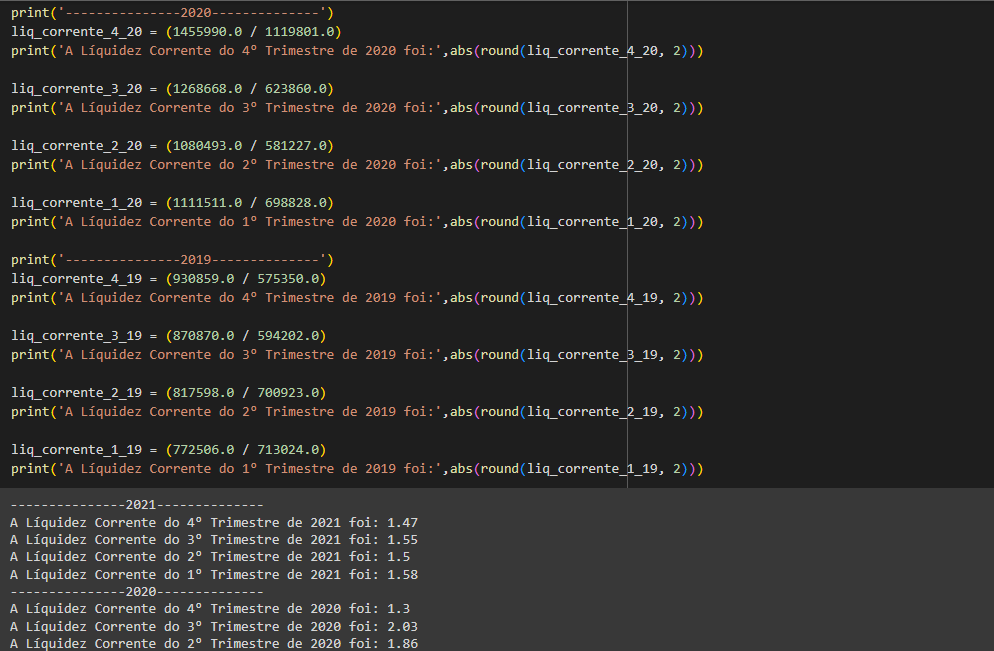
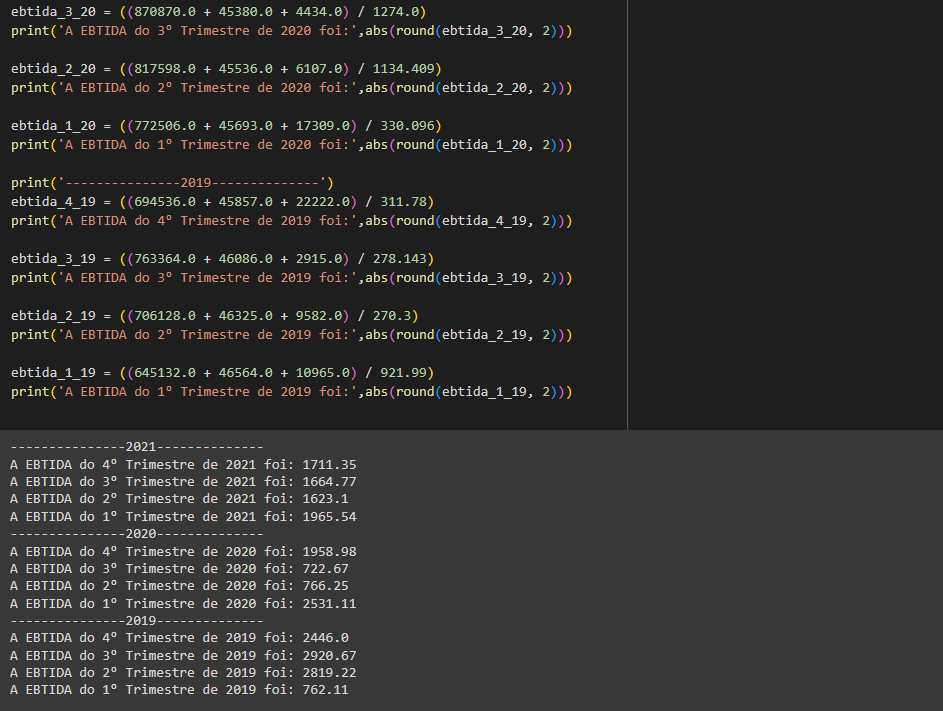
Fonte: Equipe Taurus

Figura 59 - Liquidez Corrente

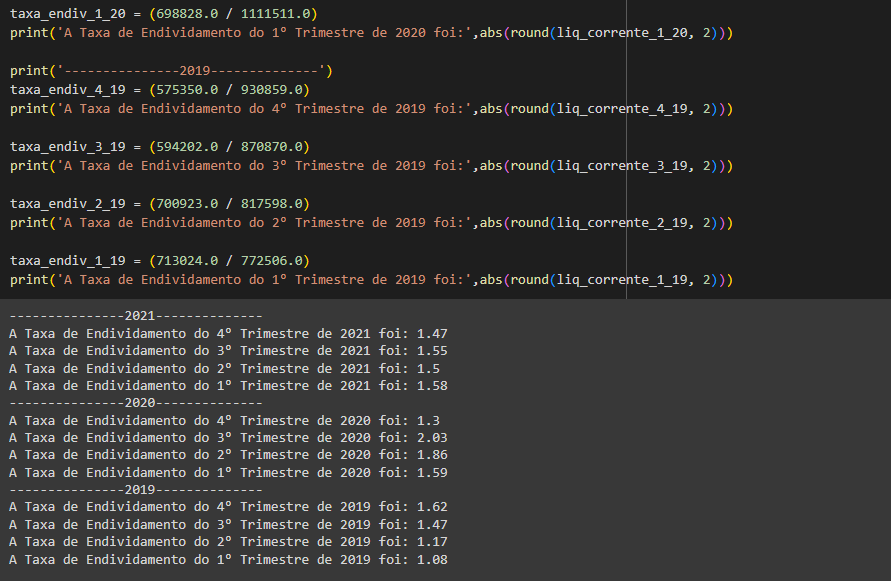


Fonte: Equipe Taurus

Figura 60 - EBTIDA



Fonte: Equipe Taurus

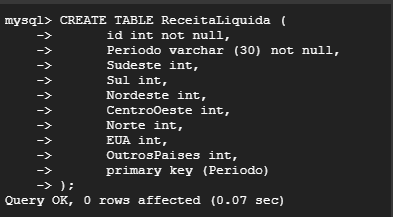
Figura 61 - Taxa de endividamento

Fonte: Equipe Taurus

## Google Cloud, MySQL e Bigquery

A equipe Rover identificou a necessidade da criação de um database manual no shell devido aos dados originários serem não estruturados(PDF).

Figura 62 - Criação da table receitaLiquida no database TaurusArmas



Fonte: Equipe Taurus

Figura 63 - Preenchendo a tabela ReceitaLiquida

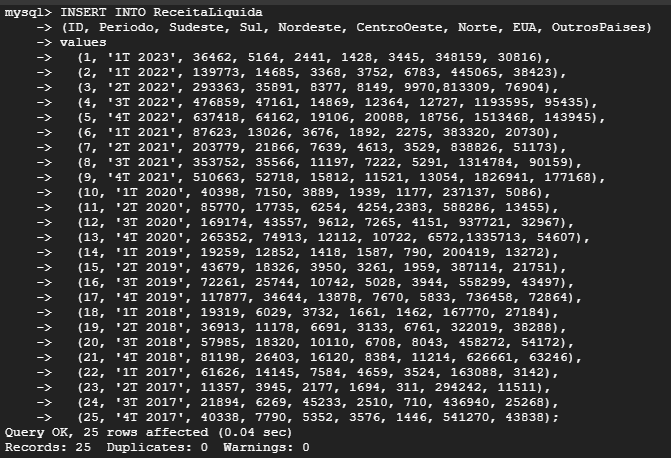
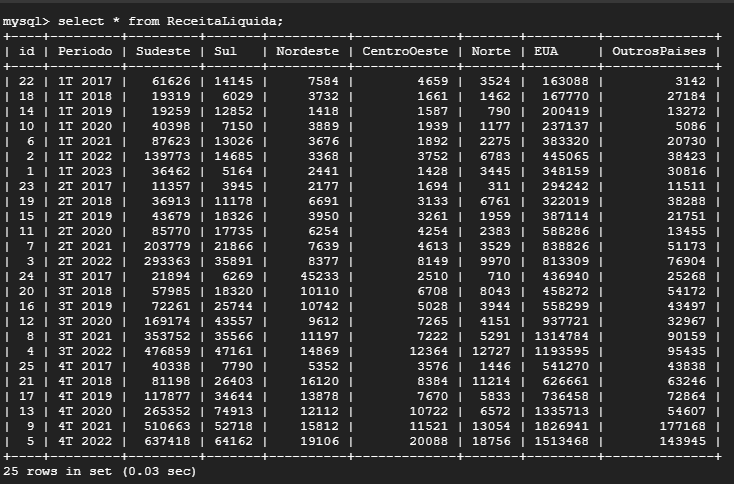
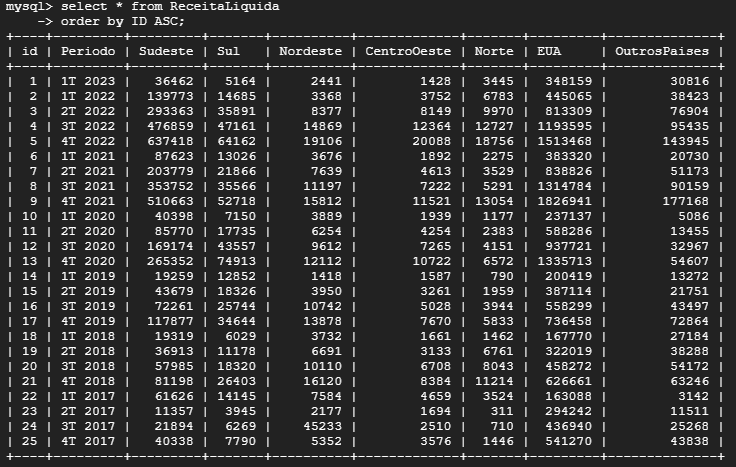


Figura 64 - Consultando se os valores foram incluídos na tabela ReceitaLiquida



Fonte: Equipe Taurus

Figura 65 - Ordenando em ordem crescente visualização dos dados



Fonte: Equipe Taurus

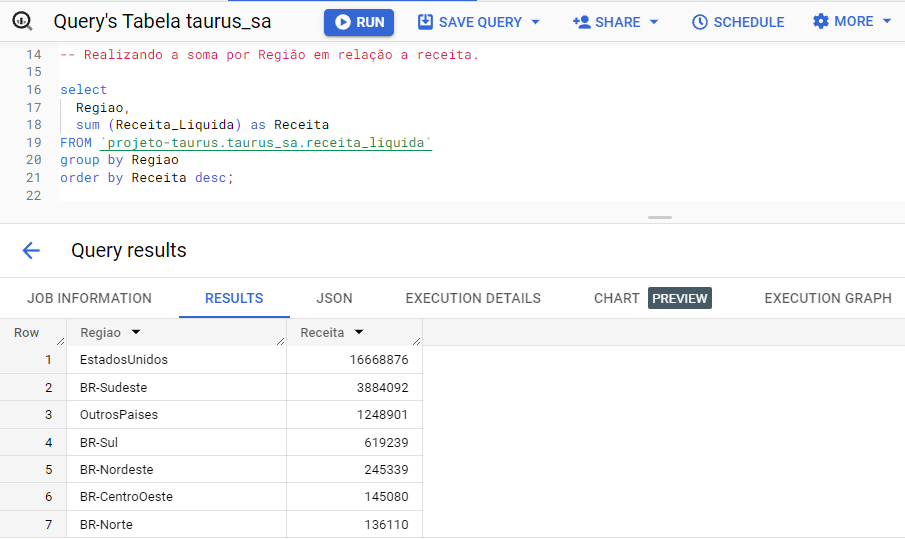
Figura 66 - Schema criado no Bigquery com SQL



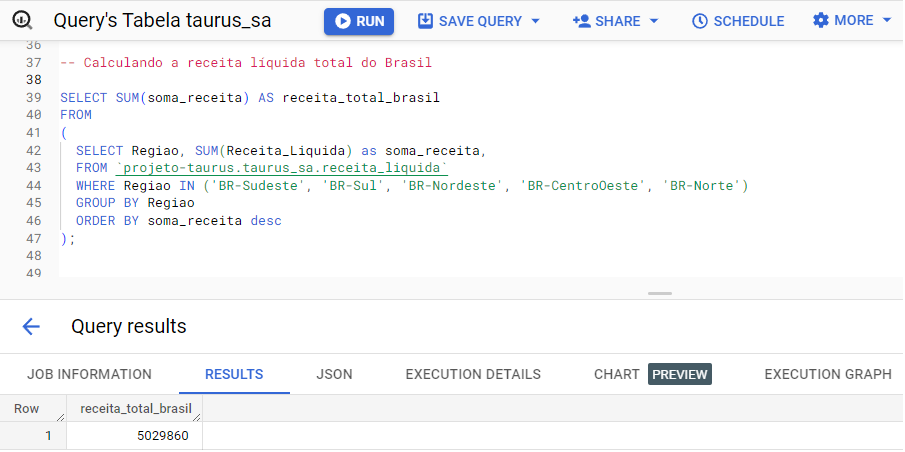
Fonte: Equipe Taurus

Figura 67 - Consultando a quantidade de linhas e valores únicos

Fonte: Equipe Taurus

Figura 68 - Soma por região em relação a receita

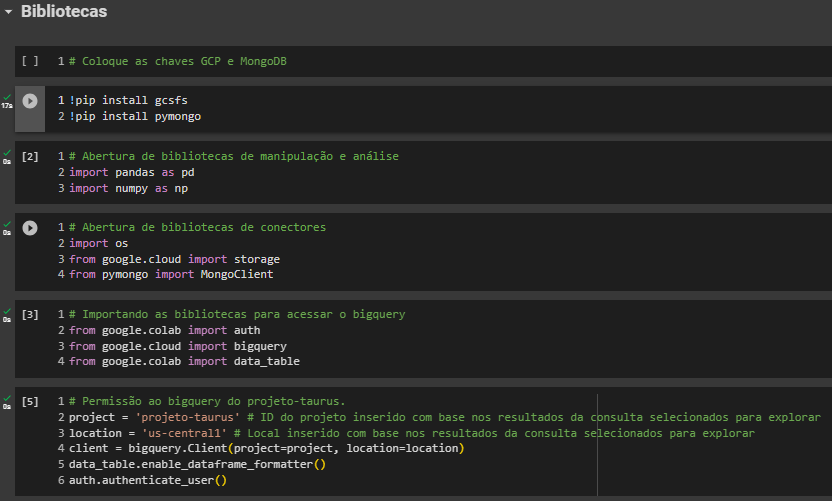
Fonte: Equipe Taurus

Figura 69 - Consulta da receita Líquida total no Brasil

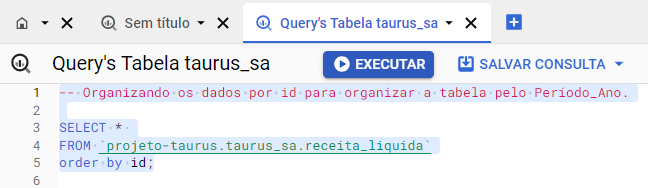
Fonte: Equipe Taurus

## MongoDB

Figura 67 - Instalação e importação das bibliotecas Cloud e bigquery e mongodb

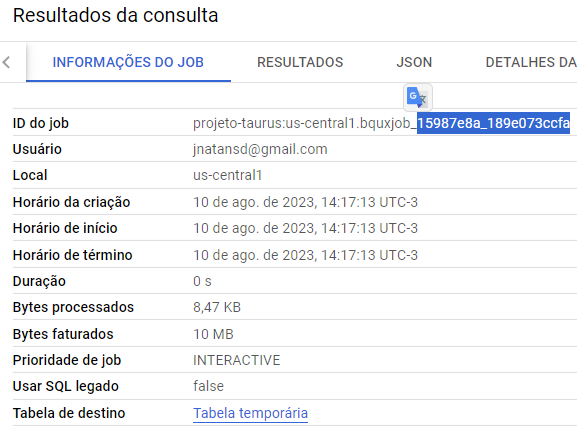


Fonte: Equipe Taurus

Figura 68 - Organizando os dados por id no Bigquery

Fonte: Equipe Taurus

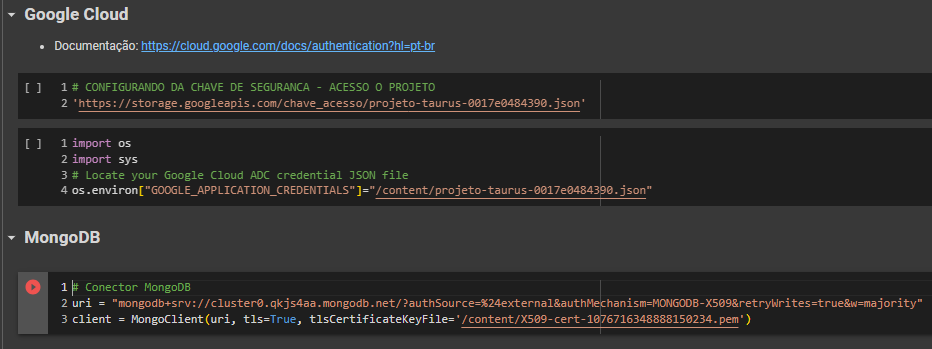
Figura 69 - Copiando o Job



Fonte: Equipe Taurus

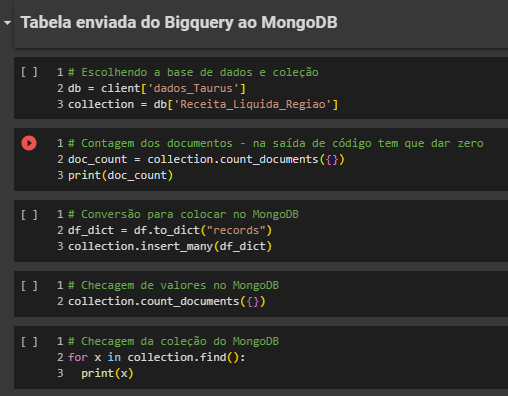
Figura 70 - Exportando a consulta do bigquery para o colab e convertendo para df pandas

Fonte: Equipe Taurus

Figura 71 - Conectando ao Mongodb

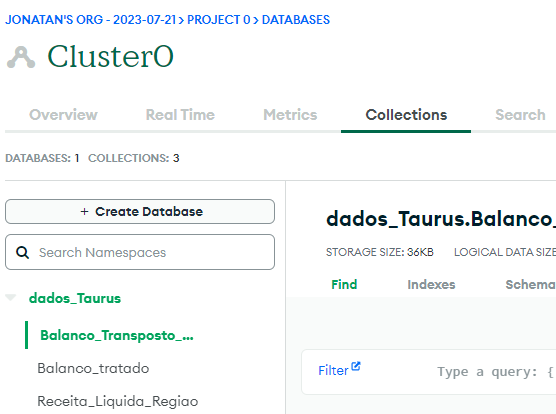
Fonte: Equipe Taurus

Figura 72 - Convertendo o formato do df para tabular e checando os valores



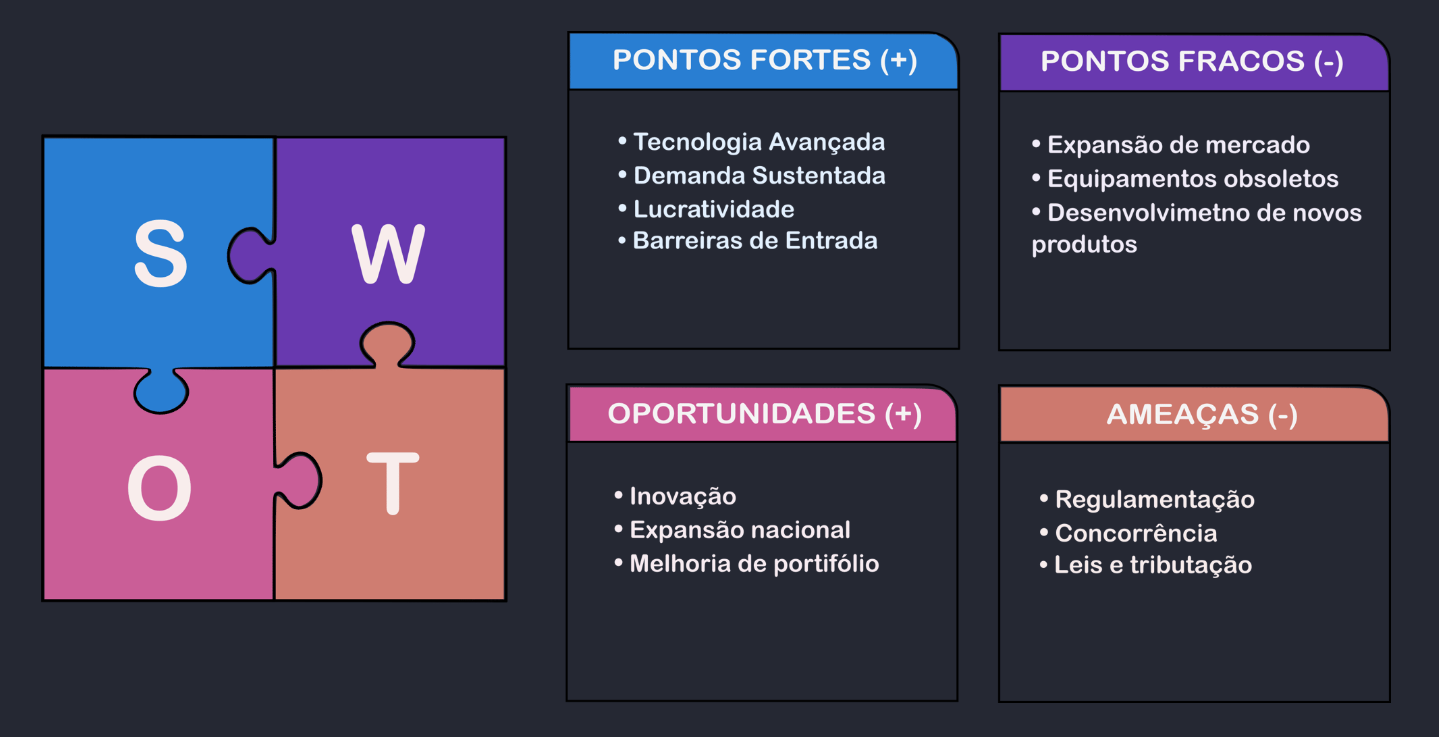
Fonte: Equipe Taurus

Figura 73 - Arquivos tratados armazenados no Mongodb



Fonte: Equipe Taurus

## Análise Swot

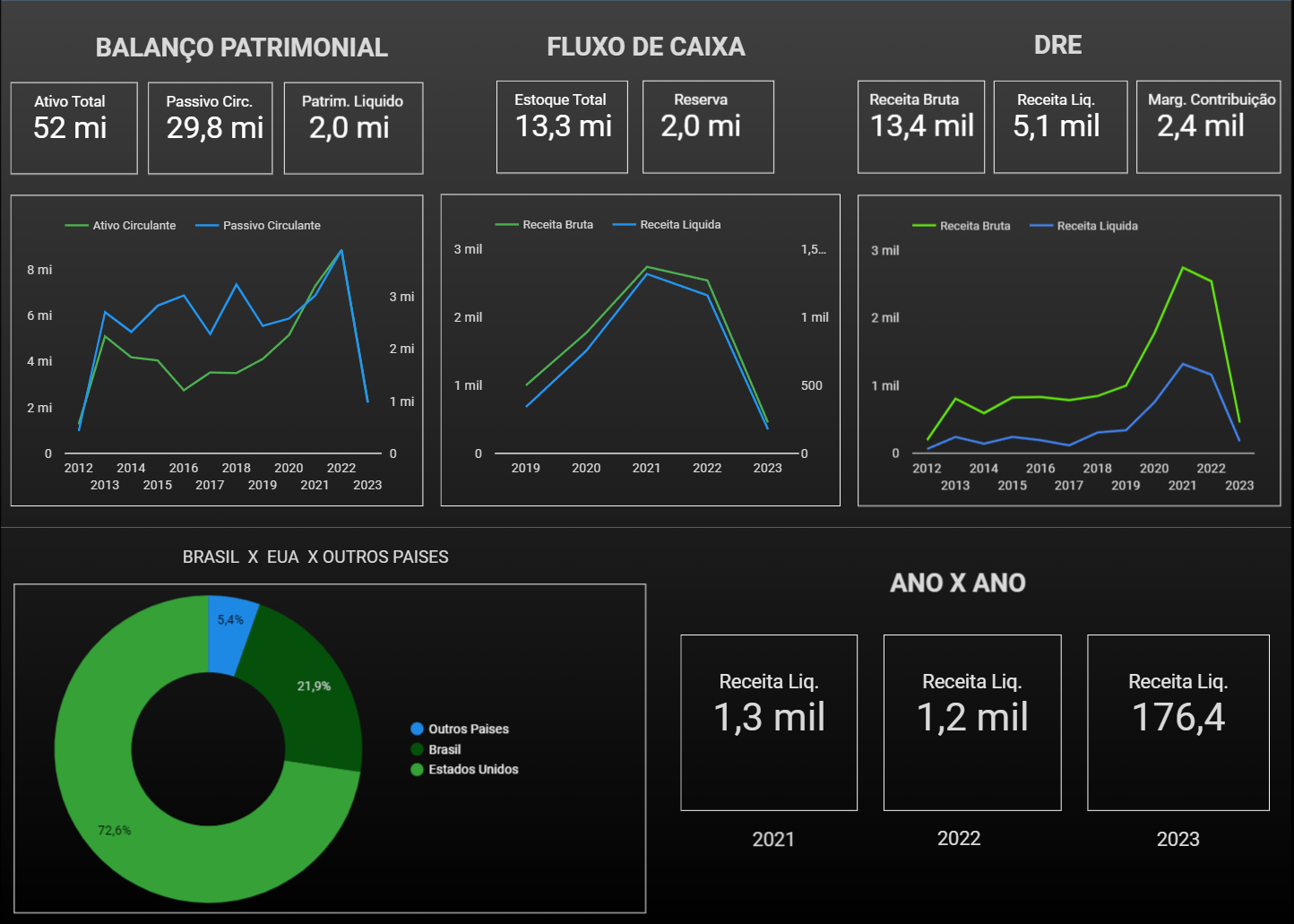


Fonte: Equipe Taurus

A equipe Rover identificou 3 grandes oportunidades através de suas análises SWOT, sendo entre elas a Expansão Nacional, uma das que podem ser melhor avaliadas após a apresentação dos dados referentes ao nível de compras por região atual no Brasil.

## 

## Looker Studio



Fonte: Equipe Taurus



Fonte: Equipe Taurus

## Links Importantes

Colab Análise de dados:

[Projeto TAURUS.ipynb](https://colab.research.google.com/drive/1t9gDHuRcXzm07N-ngMAq6fsGFwnIt5Hk?usp=sharing)

Colab da etapa MongoDB:

[Carga\_MongoDB\_Taurus\_sa.ipynb](https://colab.research.google.com/drive/1OY4liOQLVLuuLETvcgvf88SRjDQgsVl7?usp=sharing)

Bucket:

<https://console.cloud.google.com/storage/browser/taurus_armas>

Apresentação Online:

<https://gamma.app/docs/Analise-de-Dados-da-Empresa-Taurus-Armas-02-rddd3zr6axox228?mode=doc#card-h01g0bm4qvoggwf>

Dashboard: <https://lookerstudio.google.com/reporting/8f59d9ea-d339-417b-8627-a234653b408b>

Chaves necessárias:

<https://drive.google.com/drive/folders/1fdvPx507x_uCZuKHOulRHRj3TFda6Fpo?usp=sharing>

## Conclusão

Chegamos ao fim da análise da empresa Taurus, ministrada pela equipe Rover em parceria com a Soulcode. Recomendamos que, caso tenham interesse em saber um pouco mais dos bastidores de como elaboramos nosso projeto, acesse esse link: (<https://docs.google.com/document/d/1DJvjF0G5x-Z76bSR0NYjwkOmbc9MKpfhbQDcAnQhyVI/edit?usp=sharing> )

Nele constará todas as nossas etapas de brainstorming, da primeira reunião até a última, além de um pouco sobre como elaboramos algumas etapas visuais e acessíveis.