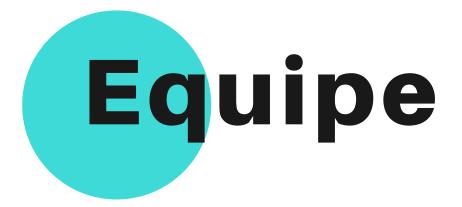


Salário dos Brasileiros

O Poder de Compra do Servidor Público da Região Sudeste







Débora Azevedo



Igor Cruz



Rodrigo Silva



Sabrina Lopes



OBJETIVO DO PROJETO

- Analisar o poder de compra do servidor público da região Sudeste
- Impacto da inflação no poder de compra, de acordo com dados sobre salário
- Poder de compra



CONJUNTO DE DADOS

- Portal de transparência das capitais
- IBGE
- IPEAData



CONTEÚDO DOS DADOS

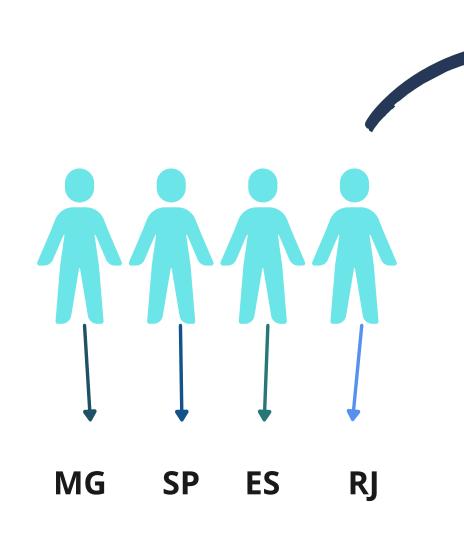
- Tabelas dos anos de 2020 e 2021
- São Paulo, Rio de Janeiro, Minas Gerais e Espírito Santo
- IPCA nacional e a nível estadual
- PIB Nacional



METODOLOGIA ÁGIL



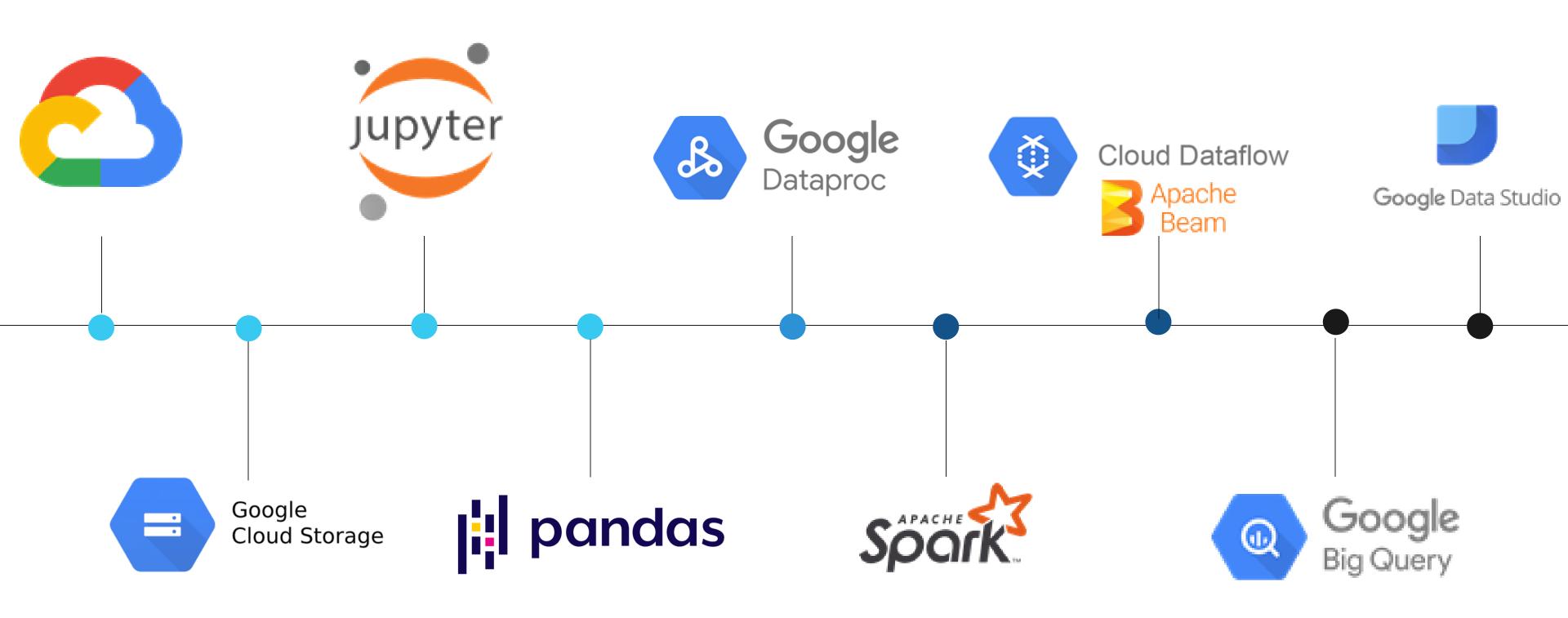
Trabalho realizado a partir de pequenos ciclos de atividades





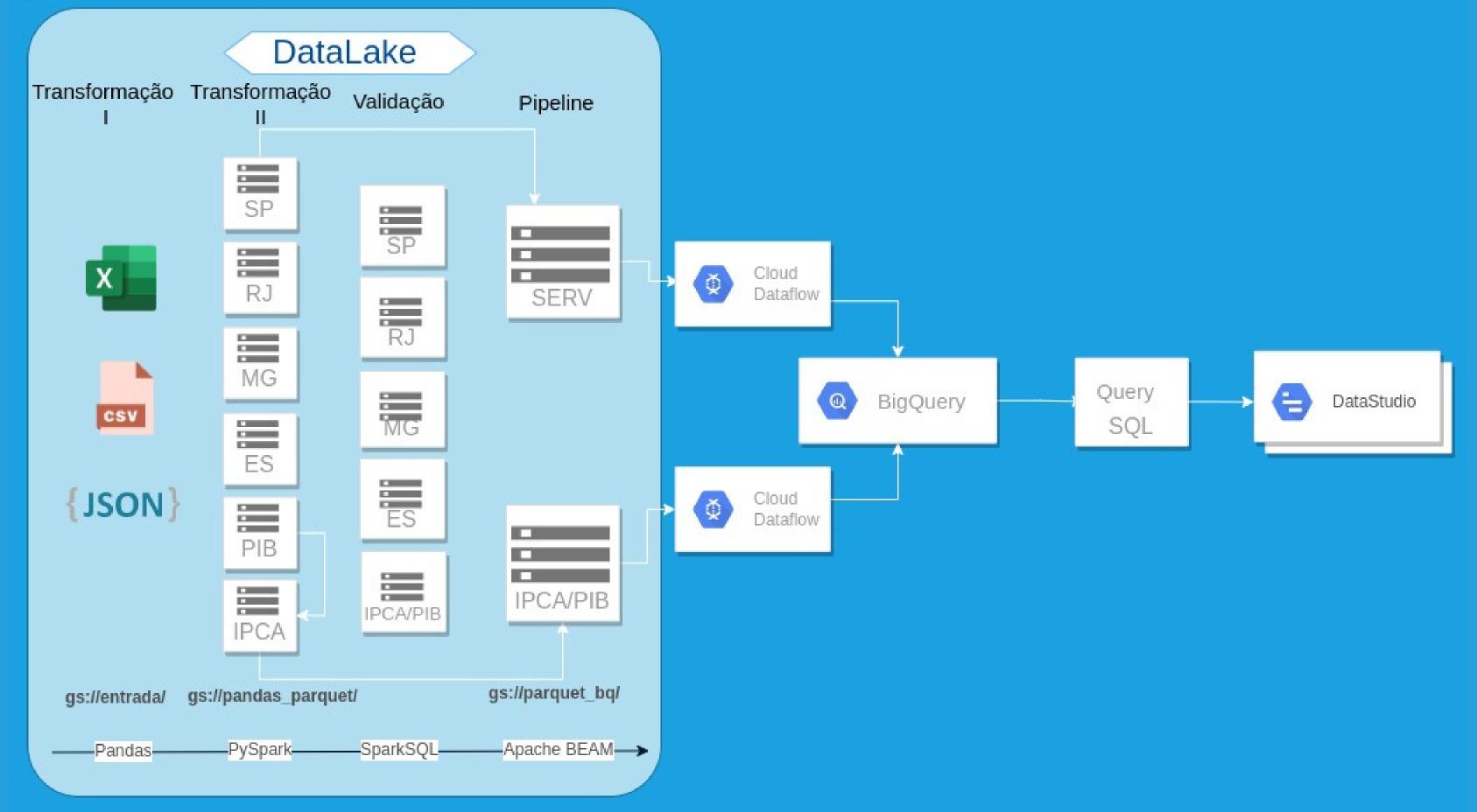


FERRAMENTAS UTILIZADAS





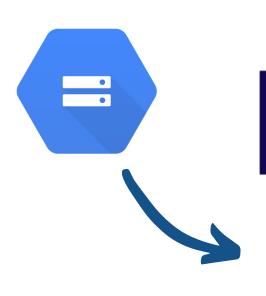
Google Cloud Platform







ETAPAS DO PROCESSO





- Extração de dados
- Concatenação
- Verificação de inconsistências
- Limpeza de dados
- Transformações
- Otimização dos dados
- Exportação dos arquivos na extensão parquet

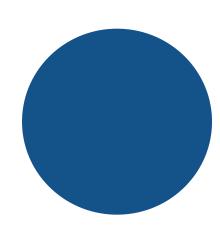


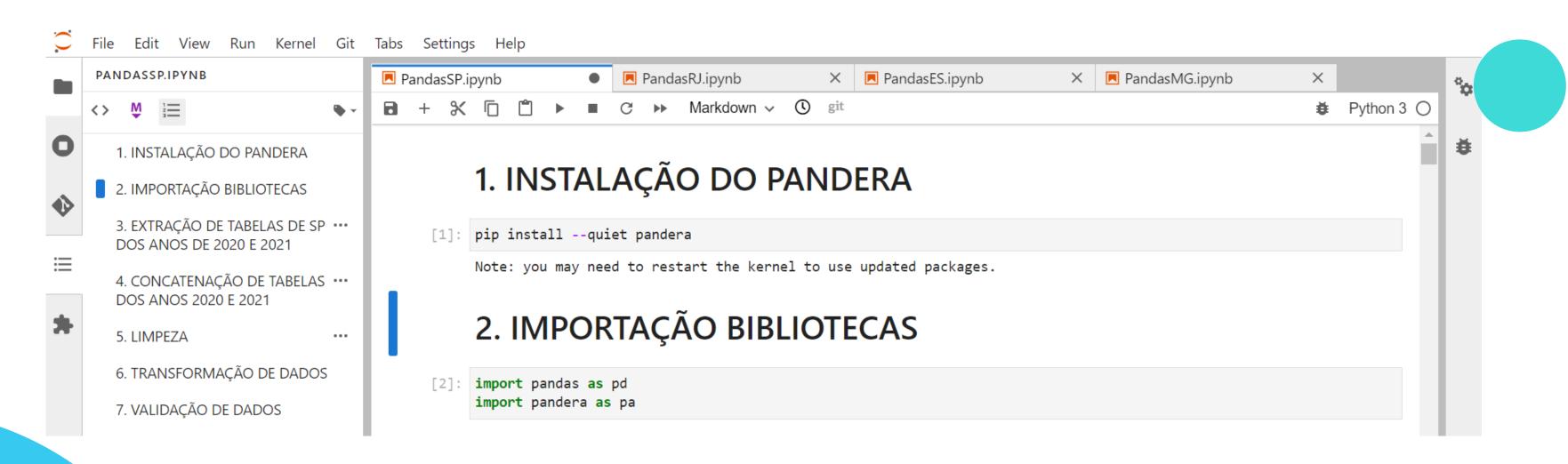
Validação dos dados



Instalação e importação de bibliotecas









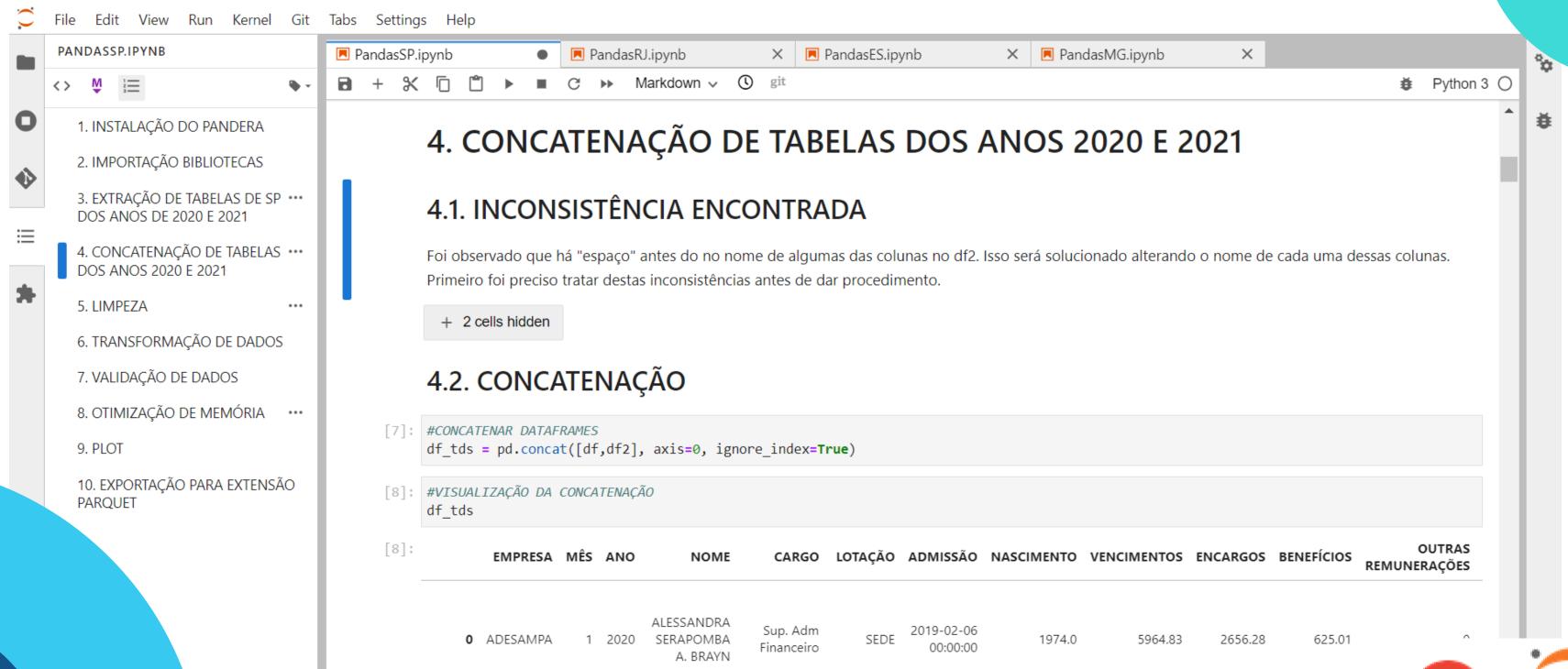


Extração de dados





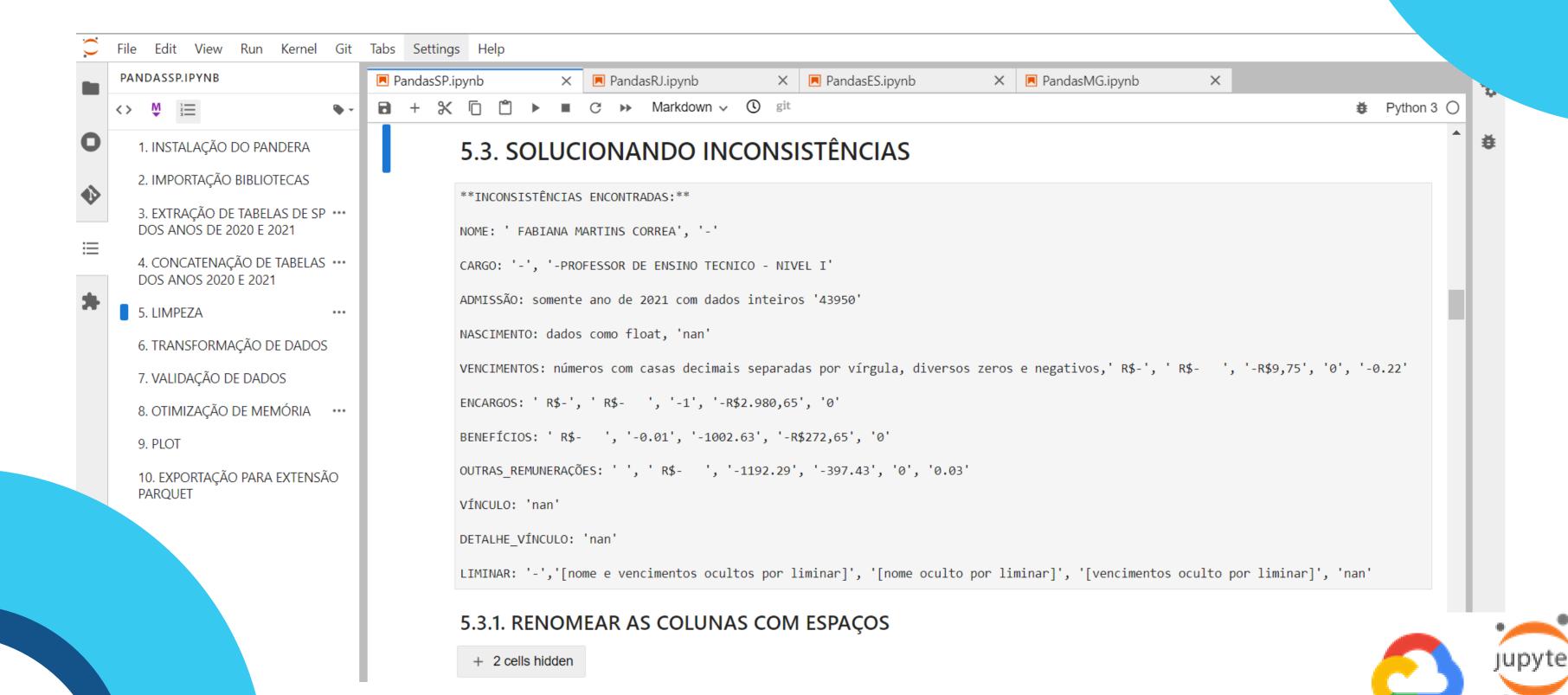
Concatenação



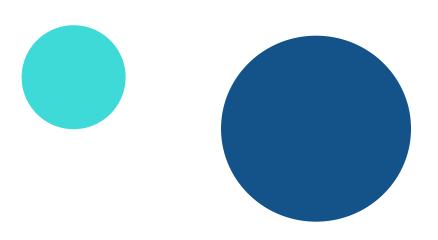




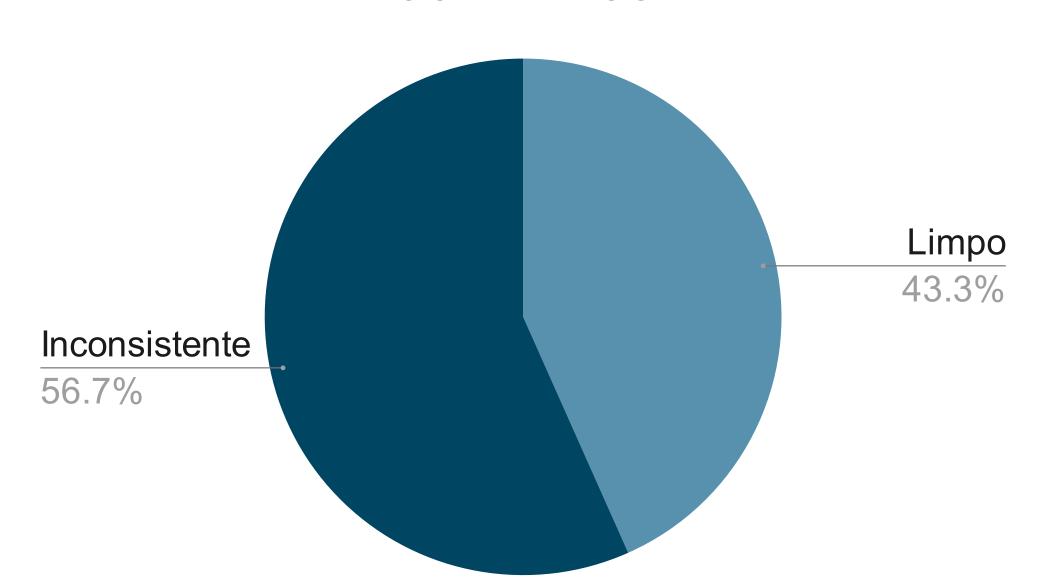
Verificação de inconsistências e Limpeza de dados







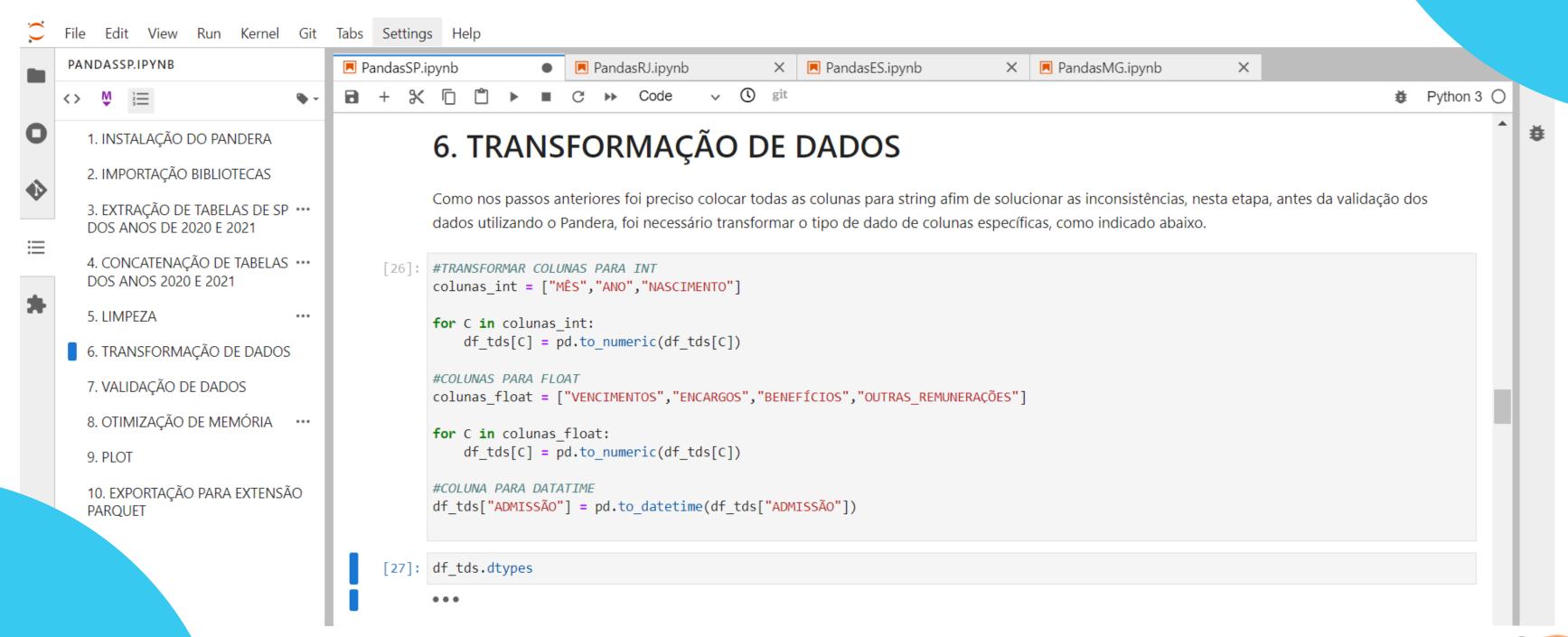
QUALIDADE DOS DADOS COLETADOS







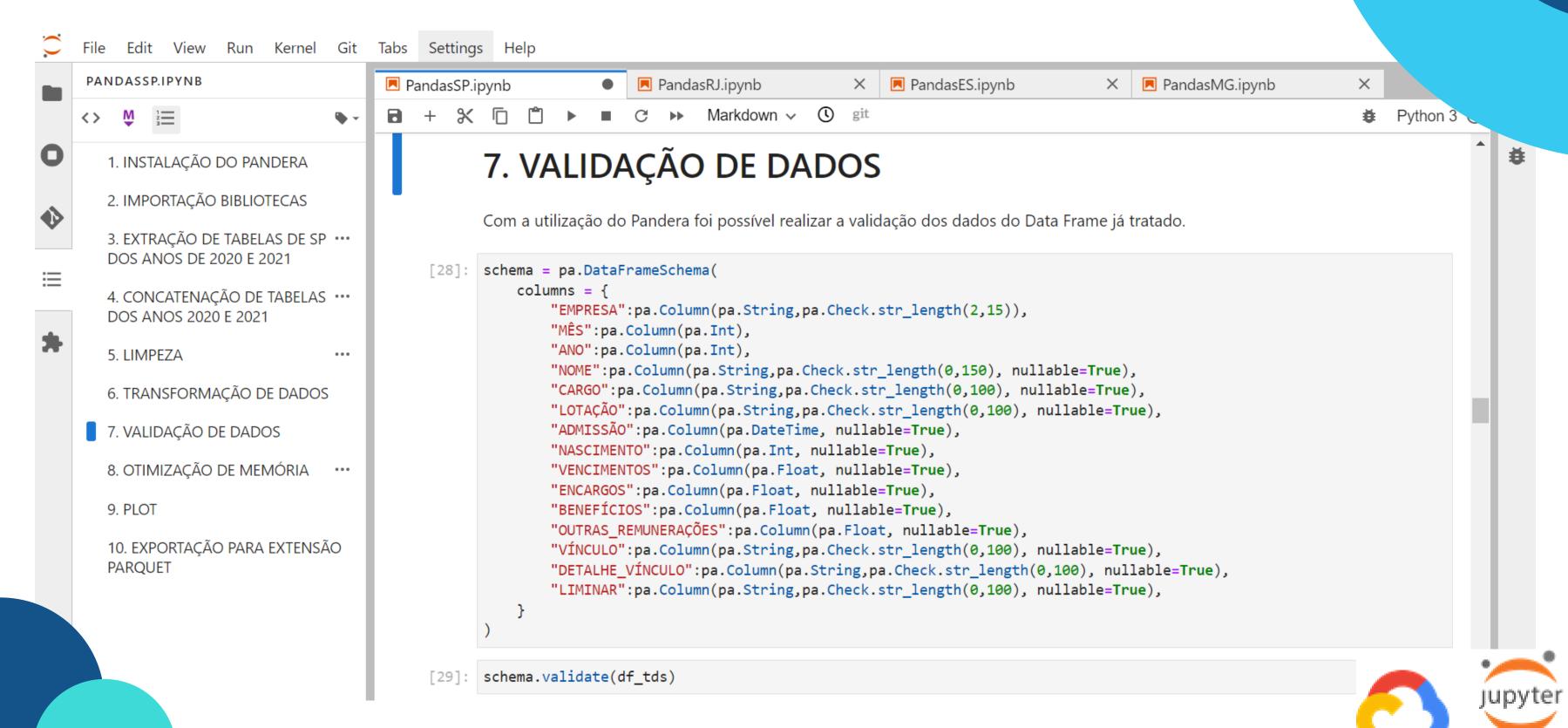
Transformações





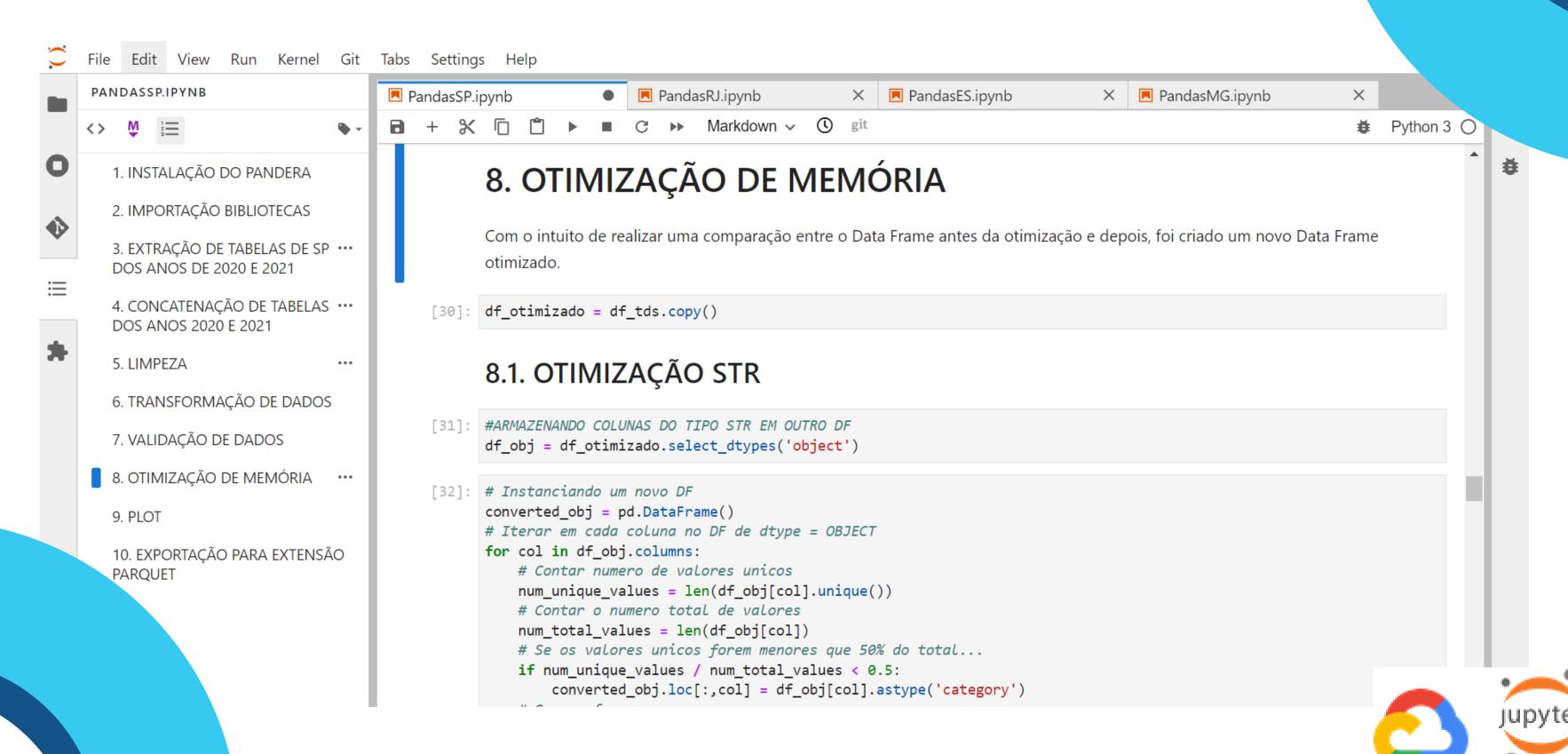


Validação dos dados

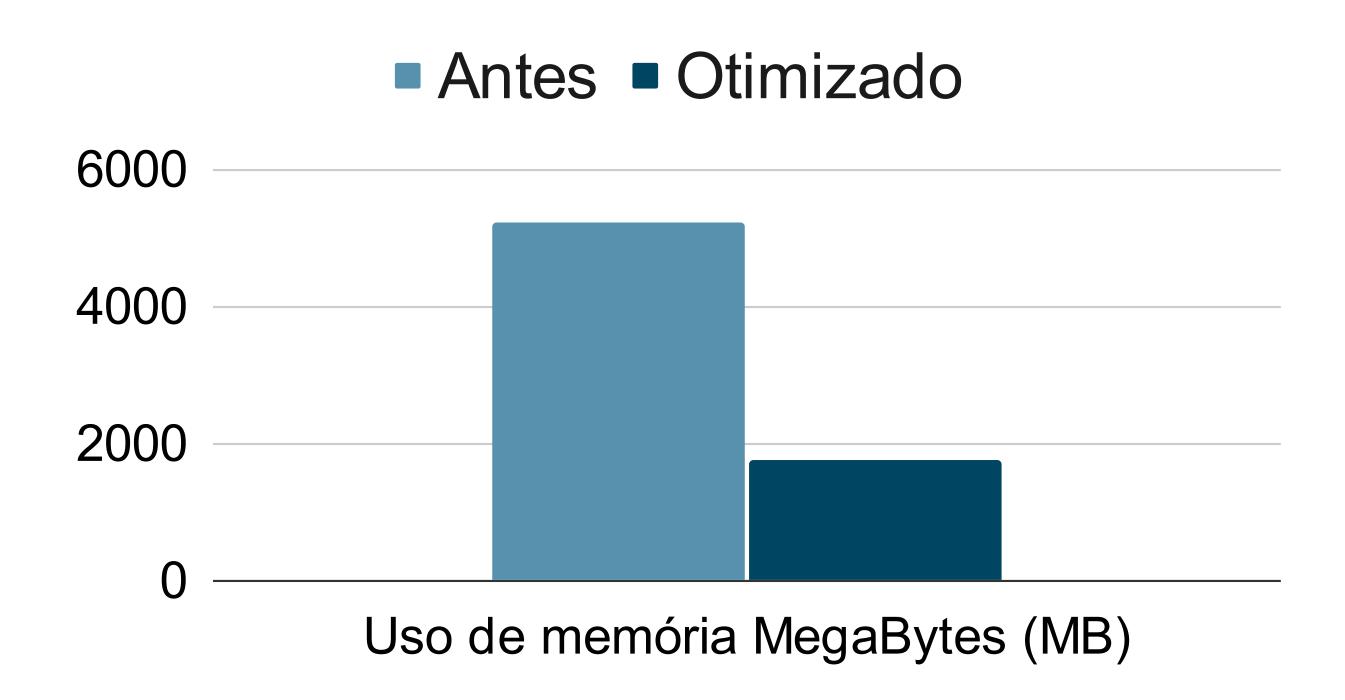




Otimização de memória







Otimização de memória





SALVAMENTO DOS ARQUIVOS



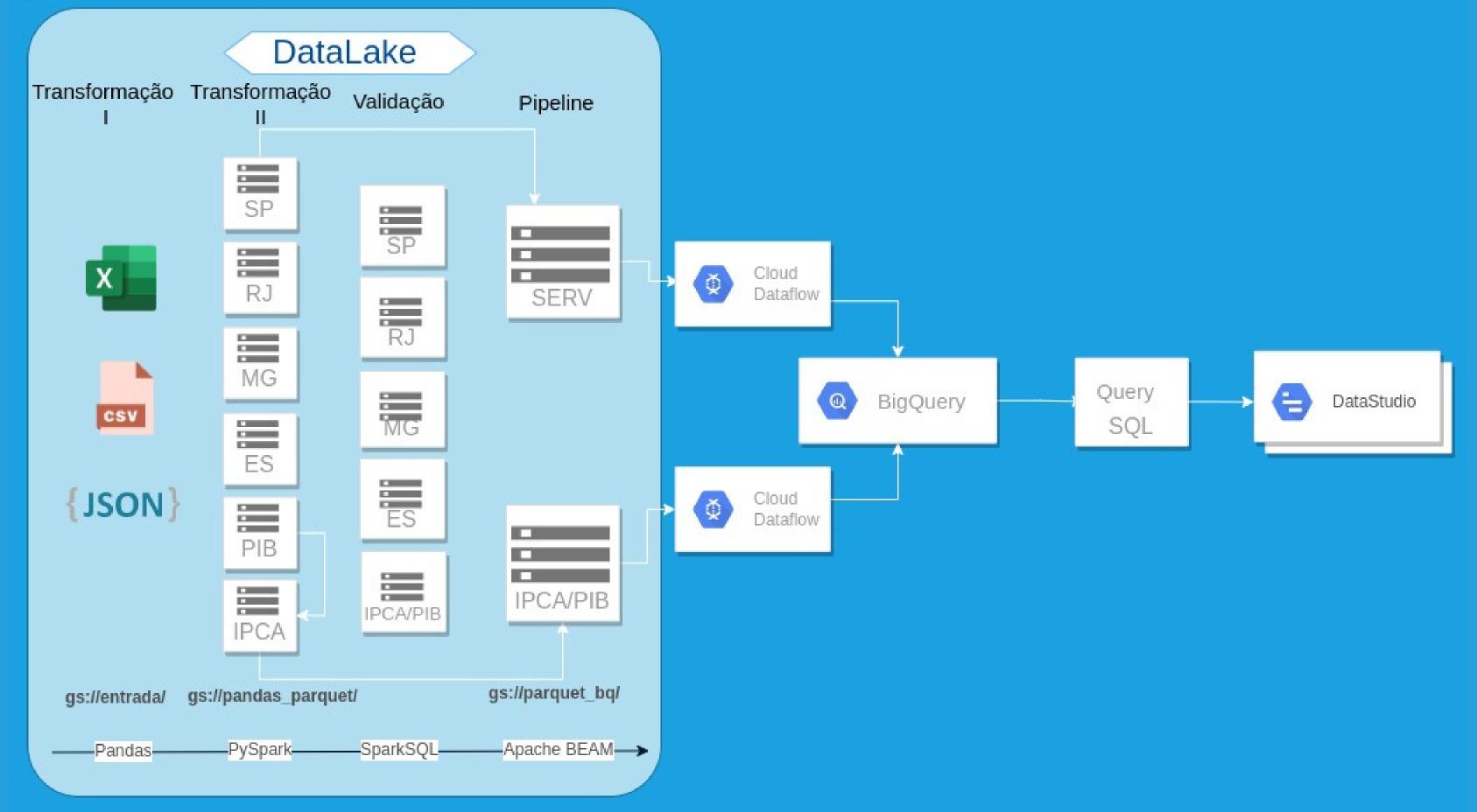
 Otimização no tempo de processamento

- Armazenamento em formato colunar
- Arquivo compactado (snappy)





Google Cloud Platform

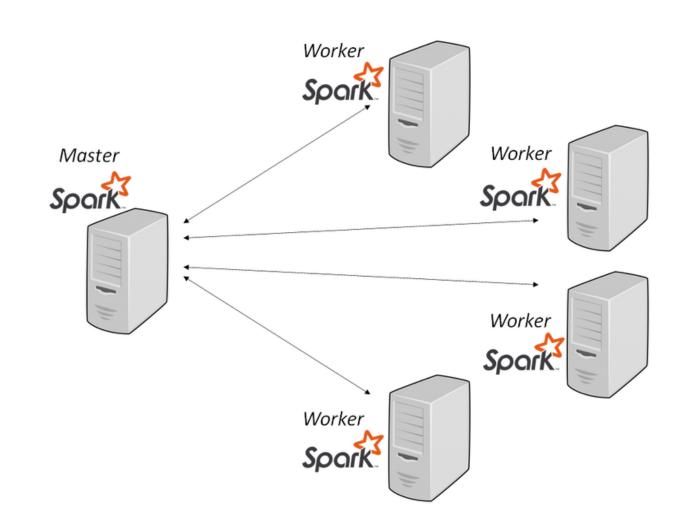




Framework opensource para processamento de dados de forma paralela e distribuida.

















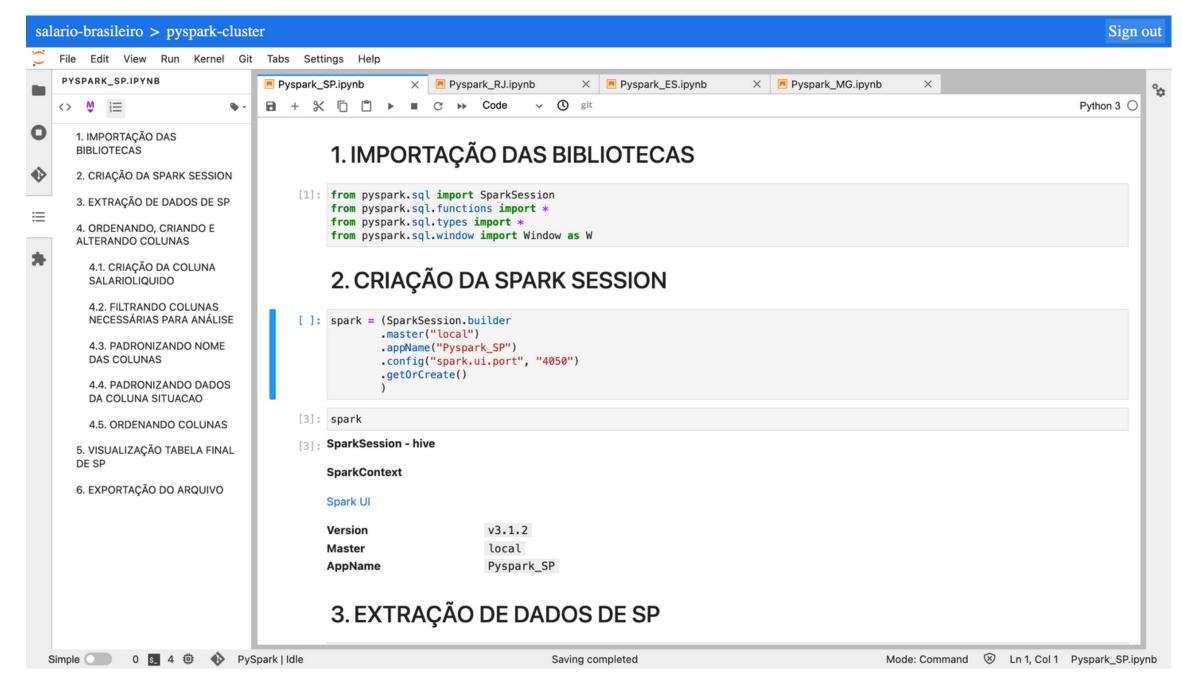


Criação de colunas

Renomeação de colunas

Ordenação colunas

Concatenação



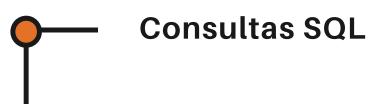






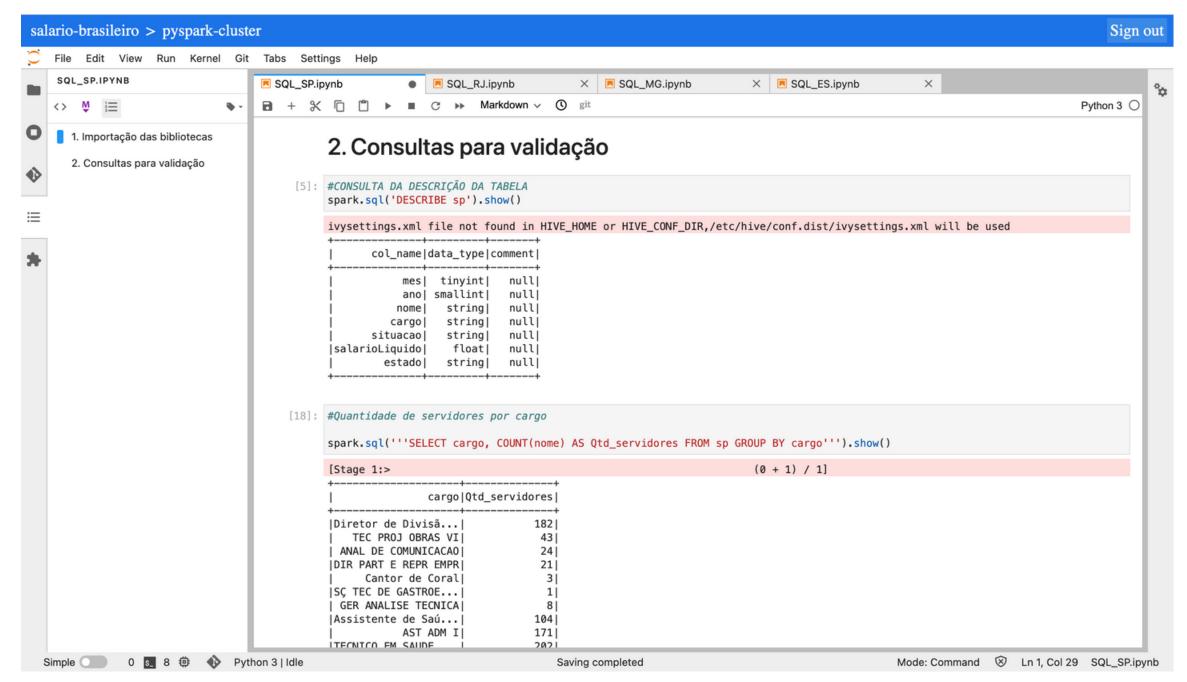






Validação do dataframe

Inconsistência nos dados!





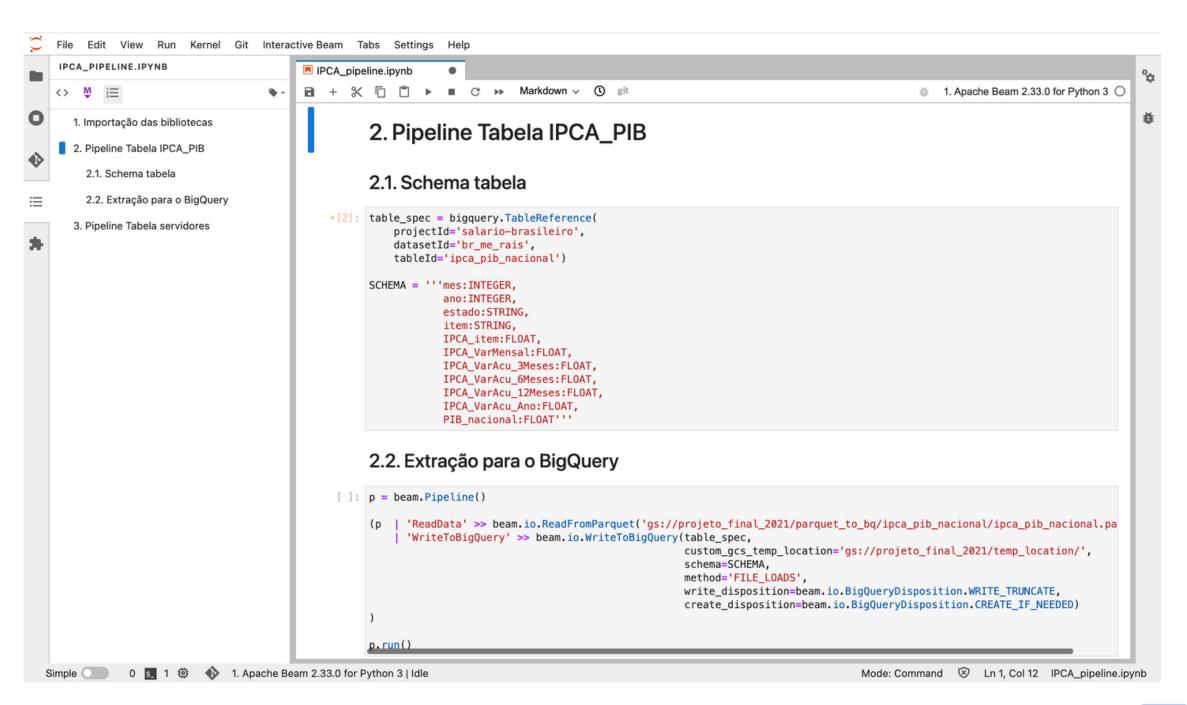






Modelo de Pipeline para tabela servidores

Extração pra o BigQuery



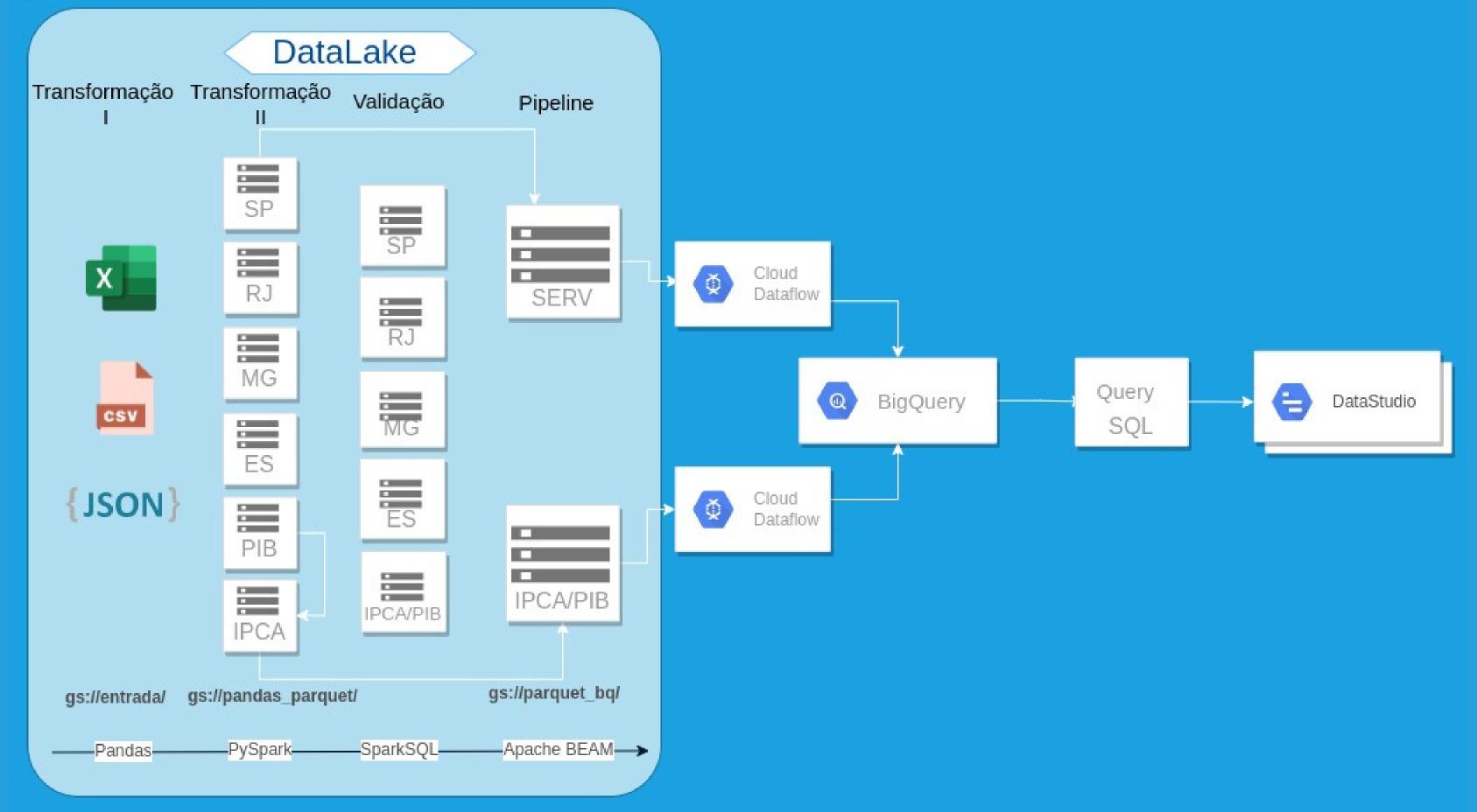








Google Cloud Platform





Como aplicar tudo isso e conectar a tecnologia ao mundo dos negócios?

Consolidamos todas as estruturas de dados da origem, num grande universo de dados



A partir dele podemos montar consultas simples, ou complexas, online ou em lote, ou até inserir dados de forma massiva

O cenário de negócio:

Impacto da inflação no salário do servidor público

Dados de IPCA

Dados de PIB

Salário dos Servidores



Geramos uma estrutura única de dados a partir da qual produzimos novos insights

Simples

```
--Média salarial por cargo

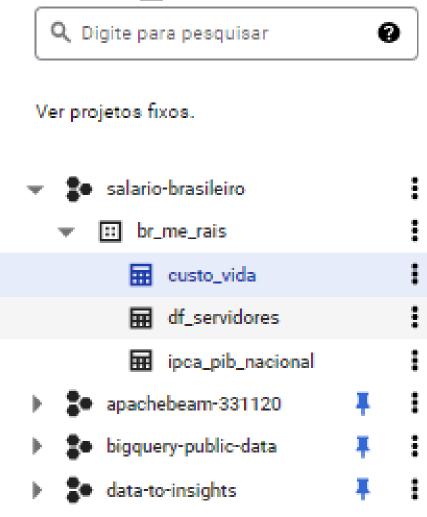
SELECT
cargo,
ROUND(AVG(salarioLiquido), 2) as media_salarial
FROM
'salario-brasileiro.br_me_rais.df_servidores'
GROUP BY
cargo

cargo
```

Médio

```
1 -- Setor com maior índice de inflação de SP de agosto de 2021
 2 SELECT
       item,
      IPCA_item
 5 FROM
       `salario-brasileiro.br_me_rais.ipca_pib_nacional`
 7 WHERE
       ano = 2021
     AND mes = 8
    AND estado = "SP"
10
     AND IPCA_item IN (
11
12
          SELECT
13
              MAX(IPCA_item)
14
          FROM
               `salario-brasileiro.br_me_rais.ipca_pib_nacional`
15
16
          WHERE
17
             ano = 2021
18
            AND mes = 8
19
             AND estado = "SP"
20
```

Complexa



```
-- ANÁLISE DO CUSTO DE VIDA DO SERVIDOR PŨBLICO DA REGIÃO SUDESTE DO BRASIL
        -- Nesta query queremos calcular a média dos salário dos servidores públicos
        -- comparados ao IPCA (inflação) e PIB (produtividade).
3
     CREATE OR REPLACE TABLE
4
        br_me_rais.custo_vida AS (
6
       SELECT
         T1.*.
8
         RANK() OVER(ORDER BY T1.media_salario_servidor DESC) AS ranking_salarios
9
       FROM (
10
         SELECT
11
           T1.ano.
12
           T1.mes.
           T1.estado.
13
14
           T2.item,
           AVG(T2.IPCA_item) AS mediaMes_IPCA_item,
15
           AVG(T2.IPCA_VarMensal) AS mediaMes_IPCA_nacional,
16
           AVG(T2.PIB_nacional) AS media_PIB,
17
           ROUND(AVG(T1.salarioLiquido), 2) AS media_salario_servidor
18
19
            salario-brasileiro.br_me_rais.df_servidores AS T1
20
21
         LEFT JOIN
            'salario-brasileiro.br_me_rais.ipca_pib_nacional' AS T2
22
23
         ON
24
           (T1.ano,
25
             T1.mes,
             T1.estado) = (T2.ano,
26
27
             T2.mes,
28
             T2.estado)
29
         WHERE
30
           T1.situacao = 'ATIVO'
31
           AND T1.situacao = 'OUTRO'
           AND T1.salarioLiquido IS NOT NULL
32
33
         GROUP BY
34
           T1.ano.
35
           T1.mes.
36
           T1.estado,
37
           T2.item ) AS T1
38
       ORDER BY
         media_salario_servidor DESC)
39
```

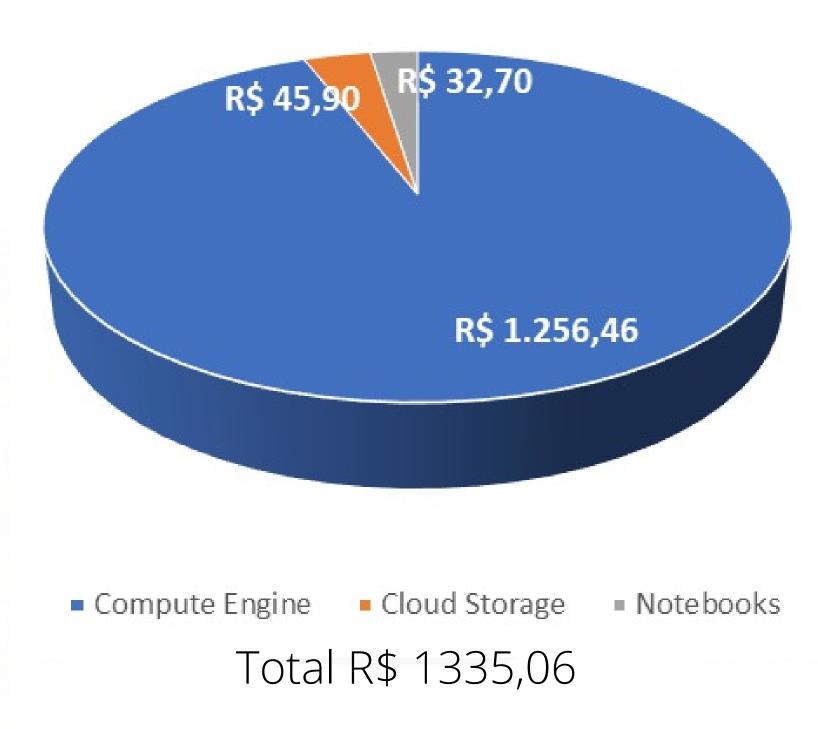


Dimensões

Time	4	pessoas
Prazo	2	semanas
Horas	96	por pessoa
Total	384	horas consumidas

Trabalhos efetuados entre 17/11 a 30/11/2021

Gastos com infraestrutura



REFERÊNCIAS

Densmore, J. Data Pipelines Pocket Reference. Moving and Processing Data for Analytics. O'Reilly 2021

Wickham, H. . (2014). Tidy Data. Journal of Statistical Software, 59(10), 1–23; Harrison, M; Petrou, T. . (2020) Pandas 1.x CookBook. Practical recipes for scientific computing, time series analysis, and exploratory data analysis using Python.

Folha de Pagamentos da Administração Indireta. Dados Abertos, 2019-2021. Desponível em:

http://dados.prefeitura.sp.gov.br/dataset/folha-de-pagamentos/. Acesso em: 23/11/2021.

Dados Abertos. Portal da transparência. Disponível em

http://transparencia.vilavelha.es.gov.br/transparenciaweb/DadosAbertos.Lista.aspx/. Acesso em: 23/11/2021.

Folha de Pagamento(Salário dos servidores). Prefeitura Miguel Pereira, portal da transparência. Disponível em:

https://transparencia.pmmp.rj.gov.br/?serv=88/. Acesso em: 23/11/2021

Tabela 7060. IBGE. Disponível em: https://sidra.ibge.gov.br/tabela/7060#resultado/. Acesso em: 23/11/2021.

IPCA. Tabela Nacional. Disponível em:

https://apisidra.ibge.gov.br/values/t/1737/n1/all/v/all/p/all/d/v63%202,v69%202,v2266%2013,v2263%202,v2264%202,v2265%202? formato=json/>. Acesso em: 23/11/2021

PIB nacional. Disponível em: http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=521274780&module=M/. Acesso em: 23/11/2021.

Agradecemos pela atenção!



Contato da Equipe



Débora Azevedo

deboramendes azevedo





Igor Cruz







Rodrigo Silva







Sabrina Lopes





