Cadastur vs CNPJs

DATA SCIENCE IMPROVES TOURISM

The Idea:

Data Base Cadastur

Base de dados Cadastur

Base de Dados Empresas Brasil - CNPJ

Data Base CNPJ

APP_cadastro_turismo

Tourism registration APP

Preparação dos dados

Modelando

Dashboard

data preparation

modeling

dashboard

Empresas a serem Cadastradas
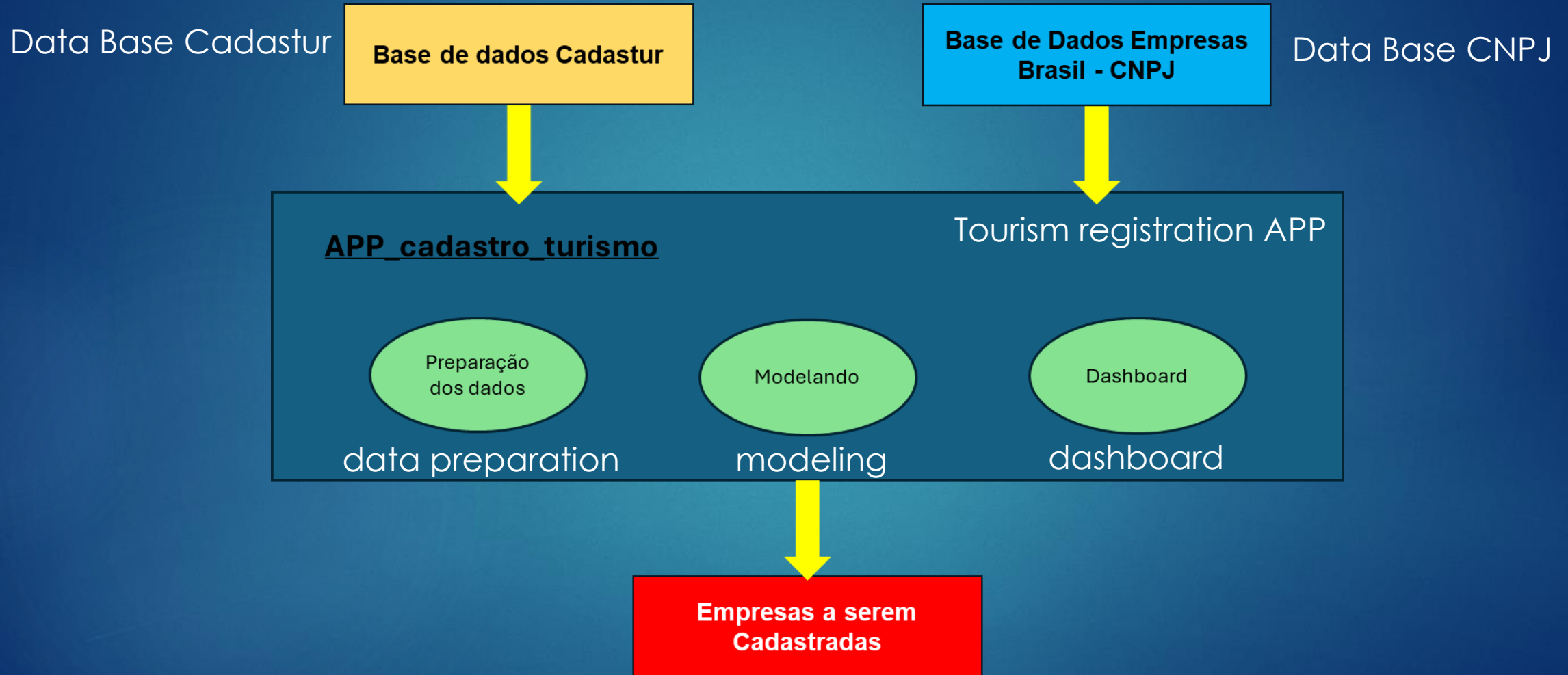
Result is a list with data on companies involved in tourism without registration in the CADASTUR

Definitions

The Idea is finding the result of crossing the two data sources, CNPJ and Cadastur, and discover what should be in Cadastur but is not.

Cadastur

CNPJ´s

Data Science

Cadastur is: The Register of Tourist Service Providers, from the Ministry of Tourism, is the register of individuals and legal entities that work in the tourism sector.

CNPJ is the acronym for "National Register of Legal Entities". This is a unique identification number assigned to each legal entity, that is, to companies and other commercial entities in Brazil. This number is used for registration and control purposes of economic and tax activities.

Two lists of two Brazilian federal Entities that have never met.
One about tourism and the other for Legal Entities.

Cadastur Original Data:

Cadastur Is federal Entity, but it has organs in the State and also Counties, to take care of the tourism in Brazil.

The companies the work with tourism are obligated to do they're on register at the Cadastur.

These registers, make the CADASTUR.

The data size is: 63,9 MB

15 entities list each one with their data characteristic.

To make a Registration: https://cadastur.turismo.gov.br/hotsite/#!/public/capa/entrar

Cadastur Original Data:

The Data at CADASTUR is open.

- CADASTUR - Tourist Camps

- CADASTUR - Travel Agencies

- CADASTUR - Entertainment Venues and Tourist

  Animation Equipment

- CADASTUR - Convention Centers

- CADASTUR - Support for Nautical Tourism or Sport

  Fishing

- CADASTUR - Entertainment and Leisure Ventures and

  Water Parks

- CADASTUR - Gastronomy

- CADASTUR - Tour Guides

- CADASTUR - Organizers of Tourist Events

- CADASTUR - Event Infrastructure Services Providers

- CADASTUR - Theme Parks

- CADASTUR - Specialized Tourism Services Providers

- CADASTUR - Tourist Transporters

- CADASTUR - Accommodation Facilities

- CADASTUR - Tourism Service Providers - Vehicle Rental

All data is here:
https://dados.gov.br/dados/conjuntos-dados/

CNPJ Original Data:

The List of "National Register of Legal Entities" - all companies in Brazil are on this register.

CNPJ is a registration issued by the Federal Revenue Service when opening a company. The CNPJ number is used to identify the business in various activities, such as issuing invoices and paying taxes.

This data collection has more than 22GB

The data need a "Dictionary of the National Register of Legal Entities", to make sense, it's totally different of the CADASTUR

Original Data: https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj

Data Understanding:

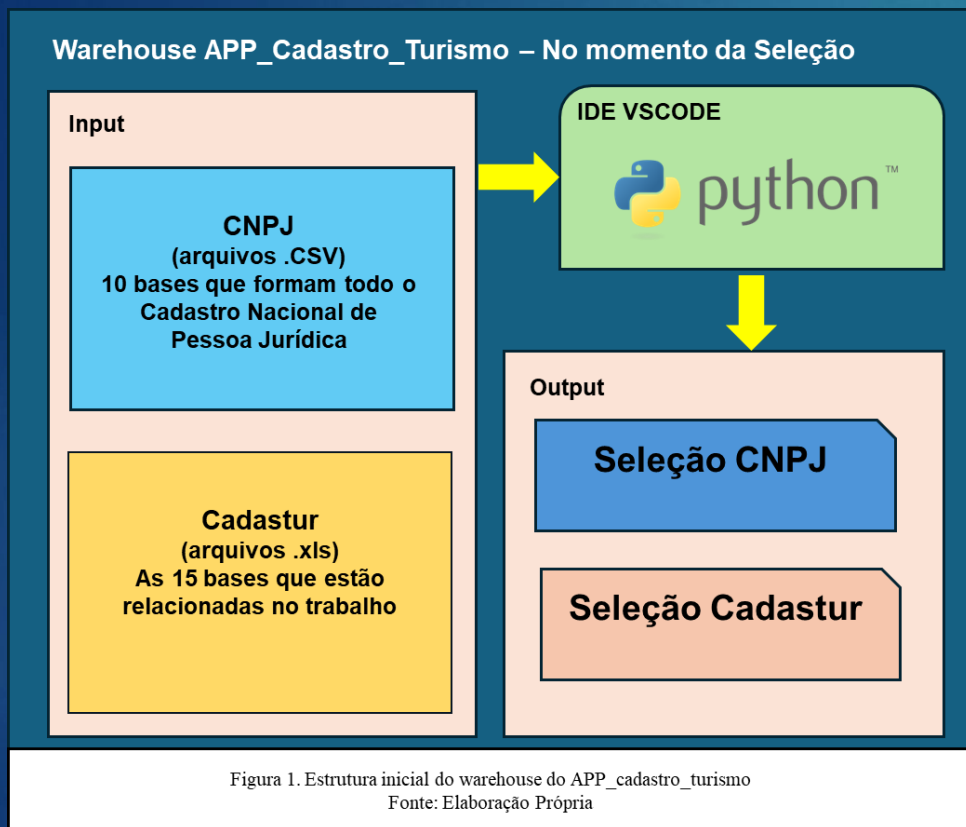| Data in all Cadasturs Bases |
|---|
| Número de Inscrição do CNPJ |
| Nome Fantasia |
| Tipo de Estabelecimento |
| Natureza Jurídica |
| Porte |
| Endereço Completo Receita Federal |
| UF |
| Município |
| CNAE(S) relacionados à atividade |

All Cadastur bases have these information

With the CNPJ base we can organize to have the same information

Important to know:

"CNAE(S) number related to the activity.
The CNAE is a National Classification of Economic Activities. It is the instrument of national standardization through codes of economic activity and framing criteria used by various tax administration bodies in the country." – Source: https://www.gov.br/empresas-e-negocios/pt-br/empreendedor/perguntas-frequentes/cadastur/o-que-e-cnae

Stage 1 – Data Preparation:

Warehouse APP_Cadastro_Turismo – No momento da Seleção

Input

CNPJ
(arquivos .CSV)
10 bases que formam todo o Cadastro Nacional de Pessoa Jurídica

IDE VSCODE

python™

Output

Seleção CNPJ

Seleção Cadastur

Cadastur
(arquivos .xls)
As 15 bases que estão relacionadas no trabalho

Figura 1. Estrutura inicial do warehouse do APP_cadastro_turismo
Fonte: Elaboração Própria

Cadastur - Step 1 Transformation:
Load data from Excel spreadsheets
Rename columns
Separate CNAE   Data Size
Format dataset   Before: 55 MB
Save as .CSV   After: 40 MB
Results:
14 .csv file

CNPJ - Step 1 Transformation:
Load the data
Rename columns
Form complete CNPJ number
Remove unnecessary columns
Format CNAE
Filter: ACTIVE companies
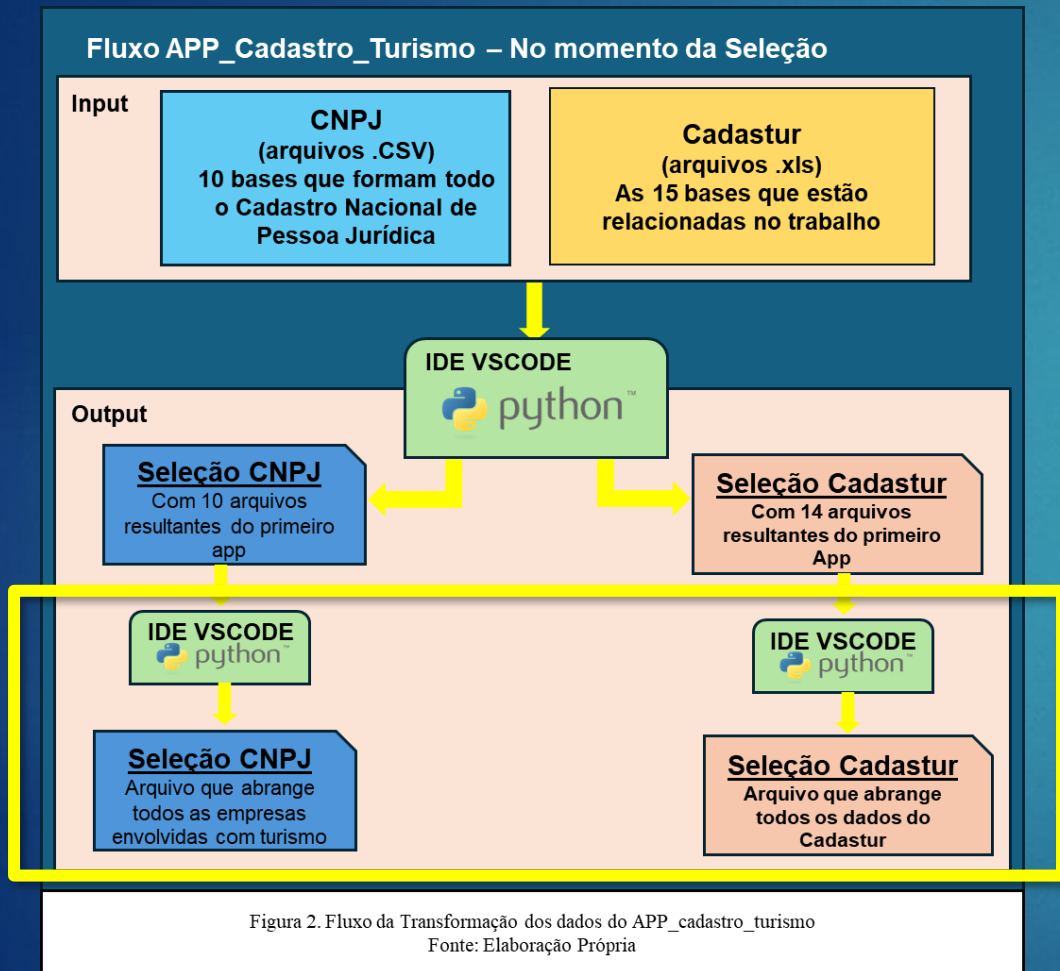Filter: CNAE Tourism
Organize Dataset
Save as .CSV
Results:
9 .csv fila

Data Size
Before: 13 GB
After: 348 MB

Stage 2 – Data Preparation:

Fluxo APP_Cadastro_Turismo – No momento da Seleção

Input

CNPJ
(arquivos .CSV)
10 bases que formam todo
o Cadastro Nacional de
Pessoa Jurídica

Cadastur
(arquivos .xls)
As 15 bases que estão
relacionadas no trabalho

IDE VSCODE
python™

Output

Seleção CNPJ
Com 10 arquivos
resultantes do primeiro
app

Seleção Cadastur
Com 14 arquivos
resultantes do primeiro
App

IDE VSCODE
python™

IDE VSCODE
python™

Seleção CNPJ
Arquivo que abrange
todos as empresas
envolvidas com turismo

Seleção Cadastur
Arquivo que abrange
todos os dados do
Cadastur

Figura 2. Fluxo da Transformação dos dados do APP_cadastro_turismo
Fonte: Elaboração Própria

Cadastur Step 2 Transformation:
Load 14 datasets
Remove unnecessary spaces
Transform 14 datasets in 1
Results:
1 .csv file with

Data Size
  Before: 40 MB
  After: 21 MB

CNPJ Step 2 Transformation:
Load 10 datasets
Remove unnecessary spaces
Transforms the municipality code into municipality
Transform 9 datasets in 1
Results:
1 .csv file with

Data Size
  Before: 348 MB
  After: 181 MB

Stage 3 – Data Preparation:

Fluxo APP_Cadastro_Turismo – No momento da

Input

CNPJ
(arquivos .CSV)
10 bases que formam todo o Cadastro Nacional de Pessoa Jurídica

Cadastur
(arquivos .xls)
As 15 bases que estão relacionadas no trabalho

IDE VSCODE
python™

Output

Seleção CNPJ
Com 10 arquivos resultantes do primeiro app

Seleção Cadastur
Com 14 arquivos resultantes do primeiro App

IDE VSCODE
python™

IDE VSCODE
python™

Seleção CNPJ
Arquivo que abrange todos as empresas envolvidas com turismo

Seleção Cadastur
Arquivo que abrange todos os dados do Cadastur

IDE VSCODE
python™

Fora_da_lista_ Cadastur.csv
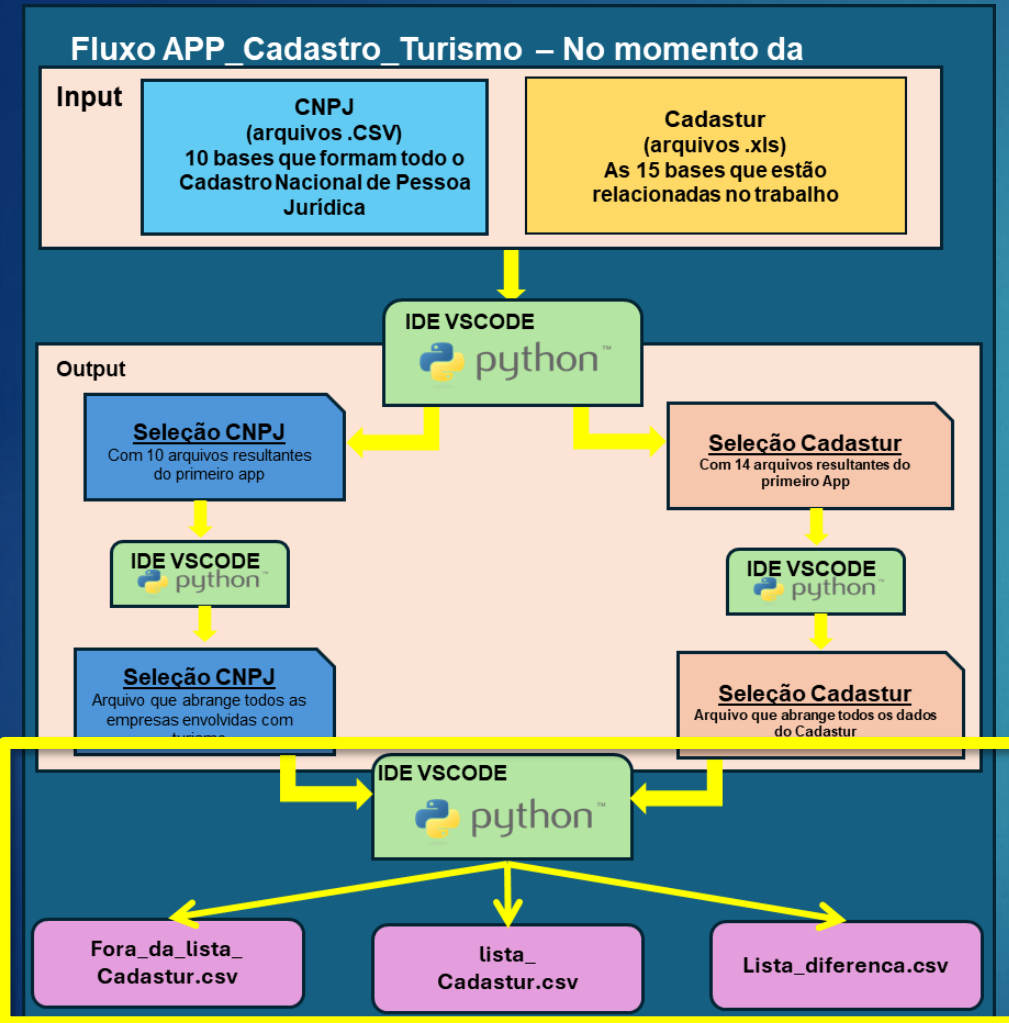
lista_ Cadastur.csv

Lista_diferenca.csv

Figura 3. Fluxo da Transformação dos dados do APP_cadastro_turismo
Fonte: Elaboração Própria

Step 3 Transformation:
Load 2 datasets
Merge 2 datasets
Select and create 3 datasets
- df_merge – all data (Cadastur data + CNPJ data)
- df_results – CNPJ in CADASTUR
- df_list – CNPJ out from CADASTUR
Transform float in integer
Transform blank spaces in zero in 'CNAE', 'CEP', 'Empresas_ID' columns
Mapping – change acronym to state names
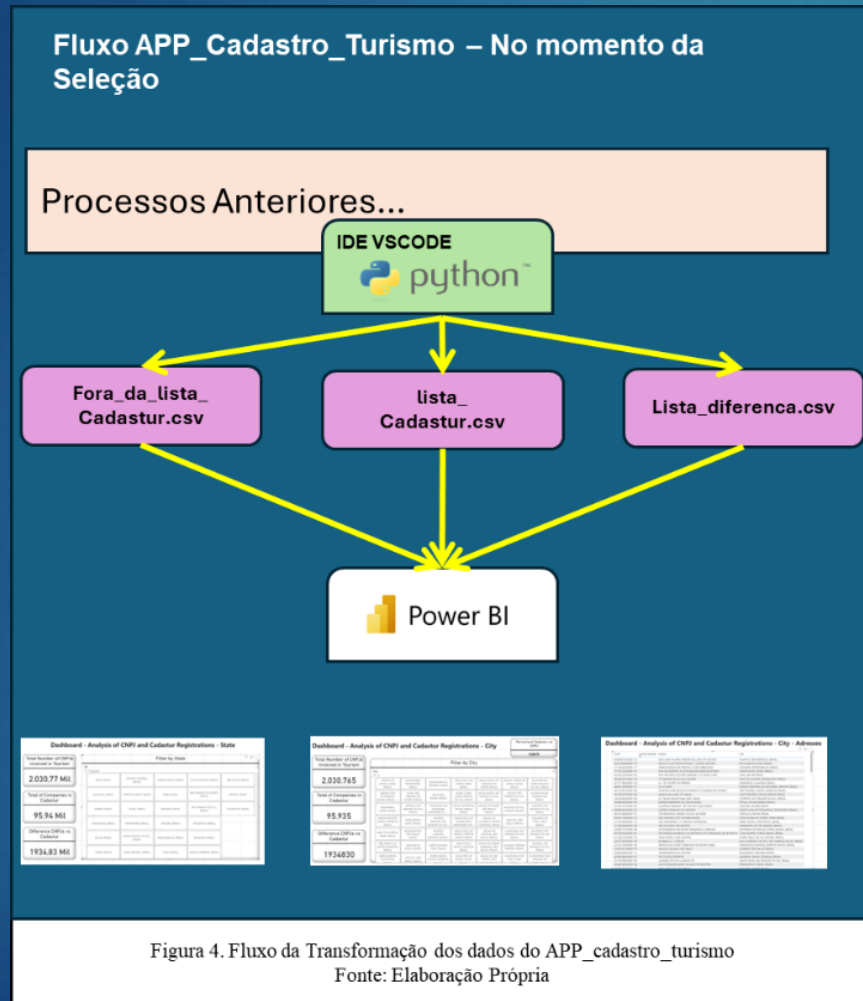Upper case configuration
Results:
3 .csv file with

Data Size
Before: 348 MB
After: 528 MB

Stage 4 – Modelling Dashboard:

Fluxo APP_Cadastro_Turismo – No momento da Seleção

Processos Anteriores...

IDE VSCODE

python

Fora_da_lista_Cadastur.csv

lista_Cadastur.csv

Lista_diferenca.csv

Power BI

Figura 4. Fluxo da Transformação dos dados do APP_cadastro_turismo
Fonte: Elaboração Própria

Step 4 Modelling & Dashboard:
Load 3 datasets
Modelling data
3 Dashboards
    State
    City
    List Addresses

Results:
3 Power BI Dashboard

Dashboard

With the three files prepared we have:

Total number of CNPJs involved in tourism by state

Total registered CNPJ in CADASTUR

Difference (companies that need to be included in the register)

Filters can be made by state

Dashboard - Analysis of CNPJ and Cadastur Registrations - State

Total Number of CNPJs involved in Tourism	Filter by State
2.030,77 Mil	UF Pesquisar

Total of Companies in Cadastur
95,94 Mil

Difference CNPJs vs Cadastur
1934,83 Mil

Filter by State UF:
ACRE, BRASIL | DISTRITO FEDERAL, BRASIL | MINAS GERAIS, BRASIL | RIO DE JANEIRO, BRASIL | SÃO PAULO, BRASIL
ALAGOAS, BRASIL | ESPÍRITO SANTO, BRASIL | PARÁ, BRASIL | RIO GRANDE DO NORTE, BRASIL | SERGIPE, BRASIL

Dashboard - Analysis of CNPJ and Cadastur Registrations - State

Total Number of CNPJs involved in Tourism
60,29 Mil

Total of Companies in Cadastur
3,28 Mil

Difference CNPJs vs
57,01 Mil

Filter by State
UF Pesquisar

ACRE, BRASIL	DISTRITO FEDERAL, BRASIL	MINAS GERAIS, BRASIL	RIO DE JANEIRO, BRASIL	SÃO PAULO, BRASIL
ALAGOAS, BRASIL	ESPÍRITO SANTO, BRASIL	PARÁ, BRASIL	RIO GRANDE DO NORTE, BRASIL	SERGIPE, BRASIL
AMAPÁ, BRASIL	GOIÁS, BRASIL	PARAÍBA, BRASIL	RIO GRANDE DO SUL, BRASIL	TOCANTINS, BRASIL
AMAZONAS, BRASIL	MARANHÃO, BRASIL	PARANÁ, BRASIL	RONDÔNIA, BRASIL	
BAHIA, BRASIL	MATO GROSSO DO SUL, BRASIL	PERNAMBUCO, BRASIL	RORAIMA, BRASIL	
CEARÁ, BRASIL	MATO GROSSO, BRASIL	PIAUÍ, BRASIL	SANTA CATARINA, BRASIL	

Dashboard

Dashboard - Analysis of CNPJ and Cadastur Registrations - City

Percentual Cadastur vs CNPJ
0,0485

percentage of companies in Cadastur - 0.0485. in the example of the city of **Cipó/BA** only 4.85%

Total Number of CNPJs involved in Tourism

103

Total of Companies in Cadastur

5

Difference CNPJs vs Cadastur

98

Filter by City

City
Cipo

CAPAO DO CIPO, RIO GRANDE DO SUL, BRASIL

CIPO, BAHIA, BRASIL

Dashboard - Analysis of CNPJ and Cadastur Registrations - City - Adresses

CNPJ	Nome_Fantasia	Adress	City
18.073.436/0001-76		05 DE AGOSTO 92 CENTRO	CIPO, BAHIA, BRASIL
30.234.522/0001-19		05 DE AGOSTO SN CENTRO	CIPO, BAHIA, BRASIL
40.666.277/0001-72		05 DE AGOSTO SN CENTRO	CIPO, BAHIA, BRASIL
27.476.903/0001-08		07 DE SETEMBRO 663 CENTRO	CIPO, BAHIA, BRASIL
19.888.218/0001-52		08 DE JULHO SN CENTRO	CIPO, BAHIA, BRASIL
23.430.368/0001-95		13 DE MAIO 35 SANTA CRUZ	CIPO, BAHIA, BRASIL
47.439.812/0001-00		13 DE MAIO SN SANTA CRUZ	CIPO, BAHIA, BRASIL
51.951.950/0001-32		7 DE SETEMBRO 0 CENTRO	CIPO, BAHIA, BRASIL
39.502.704/0001-06		7 DE SETEMBRO 296 CENTRO	CIPO, BAHIA, BRASIL
48.436.718/0001-51		AVENIDA BEIRA RIU 00 CENTRO	CIPO, BAHIA, BRASIL
22.355.393/0001-99		BOA HORA 27 CENTRO	CIPO, BAHIA, BRASIL
32.667.273/0001-80		BR 110 SN CAJUEIRO	CIPO, BAHIA, BRASIL
45.880.495/0001-29		CAJUEIRO 30 ZONA RURAL	CIPO, BAHIA, BRASIL
50.227.106/0001-09		CASA 16 PINDOBAL	CIPO, BAHIA, BRASIL
44.331.573/0001-73		CASTRO ALVES 01 CENTRO	CIPO, BAHIA, BRASIL
42.773.328/0001-90		COITE 10 ZONA RURAL	CIPO, BAHIA, BRASIL
46.667.223/0001-08		DA BANDEIRA 202 CENTRO	CIPO, BAHIA, BRASIL
46.380.032/0001-60		DA BANDEIRA S/N CENTRO	CIPO, BAHIA, BRASIL
41.835.210/0001-87		DOM PEDRO II 267 CENTRO	CIPO, BAHIA, BRASIL
12.267.105/0001-90		DOS ARGENTINOS SN CENTRO	CIPO, BAHIA, BRASIL
52.479.780/0001-06		DUQUE DE CAIXIAS 12 CENTRO	CIPO, BAHIA, BRASIL
46.700.849/0001-79		DUQUE DE CAXIAS 94 CENTRO	CIPO, BAHIA, BRASIL
05.077.013/0001-82		DUQUE DE CAXIAS SN CENTRO	CIPO, BAHIA, BRASIL
12.247.930/0001-23		ESTRADA DA RIBEIRA 76 SANTOS DUMONT	CIPO, BAHIA, BRASIL
11.891.154/0001-37		Euclides da Cunha 23 Pitomba	CIPO, BAHIA, BRASIL

Finalmente temos a lista das empresas que precisam se cadastrar no CADASTUR

Conclusion

▶ With the created system, we applied various filters for several cities in Brazil and found that a very low percentage of tourism-related companies are registered in Cadastur. For Exemple:

  ▶ São Paulo/SP: only 4% of tourism-related companies are registered in Cadastur.

  ▶ Salvador/BA: only 3.72% of tourism-related companies are registered in Cadastur.

  ▶ Rio de Janeiro/RJ: only 4.51% of tourism-related companies are registered in Cadastur.

▶ Given the alarming data about the low adoption of tourism companies in Cadastur, the implementation of the Business Location Service becomes urgent and relevant. The proposed initiative aims not only to identify and regularize non-registered businesses but also to strengthen the Brazilian tourism sector as a whole.

▶ The outlined objectives for the Business Location Service - identifying non-registered companies, supporting inspections, promoting legality, increasing security for tourists, optimizing efficiency in inspections, fostering economic development, and contributing to Cadastur transparency - are essential to overcome the challenges faced by the sector. Despite being a country full of unique attractions, beauties, rich history, and culture, Brazil receives very few visitors.

▶ it is crucial for responsible authorities to recognize the pressing need for action and mobilize efforts to encourage registration, utilizing the Business Location Service for companies that should be in Cadastur. By doing so, they will not only strengthen the tourism industry in Brazil but also ensure a safer and more transparent experience for tourists, while boosting economic development and regulatory efficiency practices.

Agradecemos!

DANIEL SANTANA - DSANTANA77@GMAIL.COM