

Optimización del talento

ABC Coorporation

Talent Focus

People Analytics.



Scrum Team

Maira - Developer

Laia - Developer

Elena - Developer

Carla - Developer y Scrum Master



Contexto

ABC Corporation es una **consultora tecnológica** estadounidense especializada en ofrecer soluciones de inteligencia artificial (IA) y aprendizaje automático (machine learning) a empresas de diversos sectores.

Necesidades:

- **Reducir la rotación de empleados**
- **Mejorar la satisfacción en el trabajo**

Estructuración del proyecto

Metodología utilizada: filosofía **Agile** con marco de trabajo **Scrum**

Sprint 1

- Análisis Exploratorio de Datos (EDA)
- Transformación
- Creación BBDD* e inserción de datos

Sprint 2

- Análisis y A/B testing
- Extracción, Transformación y Carga de datos (ETL)
- Reporte de resultados

*BBDD: Base de Datos



Estructuración del proyecto

Metodología utilizada: filosofía **Agile** con marco de trabajo **Scrum**

Sprint 1

- Análisis Exploratorio de Datos (EDA) ✓
- Transformación
- Creación BBDD* e inserción de datos

Sprint 2

- Análisis y A/B testing
- Extracción, Transformación y Carga de datos (ETL)
- Reporte de resultados

*BBDD: Base de Datos



Estructuración del proyecto

Metodología utilizada: filosofía **Agile** con marco de trabajo **Scrum**

Sprint 1

- Análisis Exploratorio de Datos (EDA) ✓
- Transformación ✓
- Creación BBDD* e inserción de datos

Sprint 2

- Análisis y A/B testing
- Extracción, Transformación y Carga de datos (ETL)
- Reporte de resultados

*BBDD: Base de Datos



Estructuración del proyecto

Metodología utilizada: filosofía **Agile** con marco de trabajo **Scrum**

Sprint 1

- Análisis Exploratorio de Datos (EDA) ✓
- Transformación ✓
- Creación BBDD* e inserción de datos ✓

Sprint 2

- Análisis y A/B testing
- Extracción, Transformación y Carga de datos (ETL)
- Reporte de resultados

*BBDD: Base de Datos



Estructuración del proyecto

Metodología utilizada: filosofía **Agile** con marco de trabajo **Scrum**

Sprint 1

- Análisis Exploratorio de Datos (EDA) ✓
- Transformación ✓
- Creación BBDD* e inserción de datos ✓

Sprint 2

- Análisis y A/B testing ✓
- Extracción, Transformación y Carga de datos (ETL)
- Reporte de resultados

*BBDD: Base de Datos



Estructuración del proyecto

Metodología utilizada: filosofía **Agile** con marco de trabajo **Scrum**

Sprint 1

- Análisis Exploratorio de Datos (EDA) ✓
- Transformación ✓
- Creación BBDD* e inserción de datos ✓

Sprint 2

- Análisis y A/B testing ✓
- Extracción, Transformación y Carga de datos (ETL) ✓
- Reporte de resultados

*BBDD: Base de Datos



Estructuración del proyecto

Metodología utilizada: filosofía **Agile** con marco de trabajo **Scrum**

Sprint 1

- Análisis Exploratorio de Datos (EDA) ✓
- Transformación ✓
- Creación BBDD* e inserción de datos ✓

Sprint 2

- Análisis y A/B testing ✓
- Extracción, Transformación y Carga de datos (ETL) ✓
- Reporte de resultados ✓

*BBDD: Base de Datos



Desarrollo del proyecto

Desafíos

- Toma de decisiones a nivel equipo
- EDA, limpieza de datos y gestión de valores nulos
- Planificación, organización del trabajo y gestión del tiempo

Hitos

- Creación de **Clases**
- **EDA** super exhaustivo
- Repositorio 100% en **inglés**
- Alcanzamos conclusiones de valor para el cliente



DEMO

- 01.** Análisis Exploratorio de los Datos (EDA)
- 02.** Extracción, Transformación y Carga de los datos (ETL)
- 03.** Análisis A/B, informe de resultados y sugerencias

Análisis Exploratorio de los Datos (EDA)

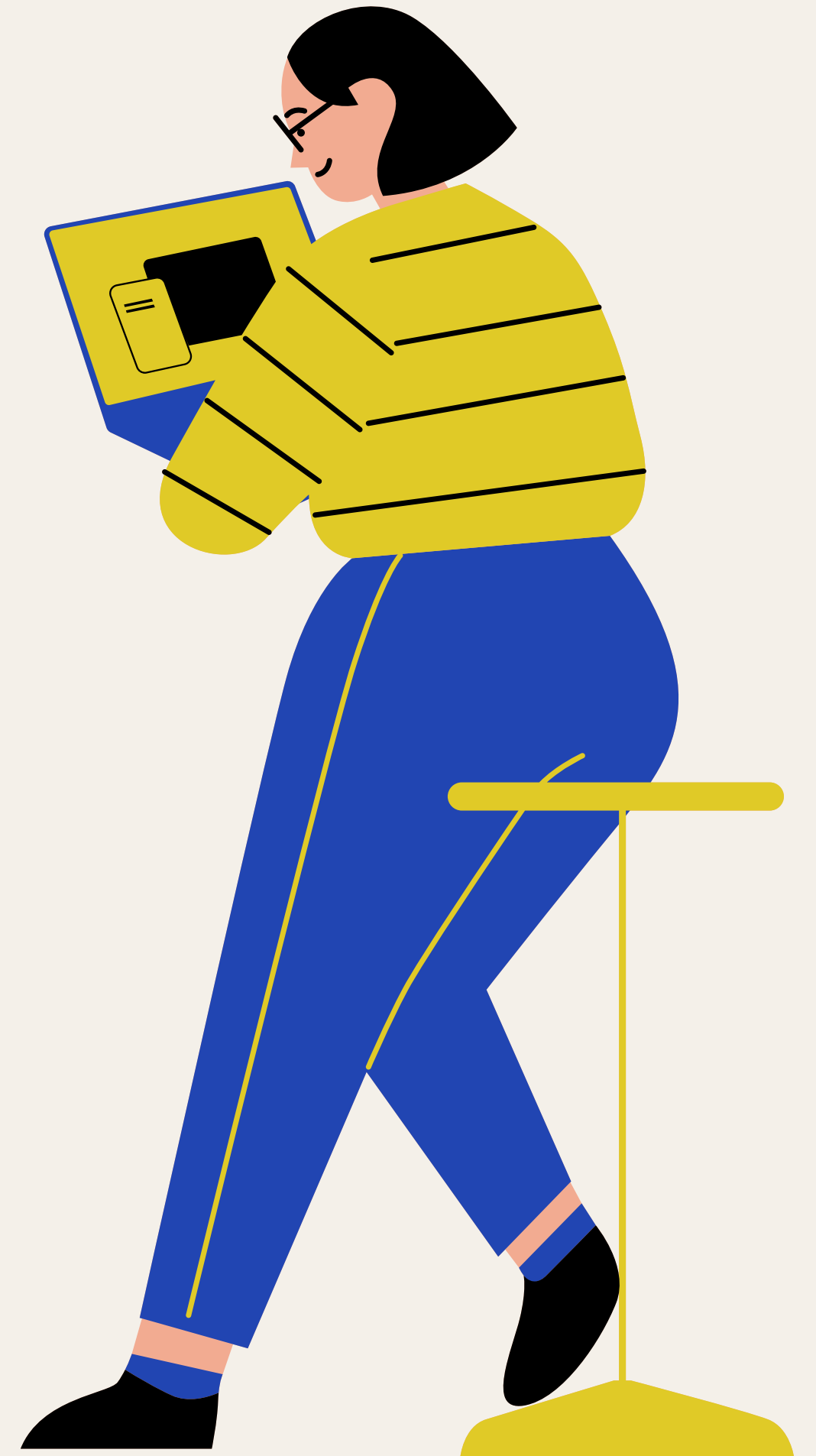
Exploración Inicial

Visualización de Datos

Análisis de Correlaciones

Análisis de Variables Categóricas

Conclusiones Preliminares



jupyter_notebooks > exploratory_data_analysis.ipynb > eda.explo_df(df, ['attrition', 'department'])

+ Code + Markdown | ▶ Run All ⌂ Restart ≡ Clear All Outputs | 📄 Variables ≡ Outline ...

base (Python 3.11.9)

29 years_since_last_promotion 1614 non-null int64

30 years_with_curr_manager 1614 non-null int64

31 date_birth 1614 non-null int64

32 remote_work 1614 non-null object

dtypes: float64(7), int64(16), object(10)

memory usage: 428.7+ KB

...

None

...

First 10 rows of the DataFrame:

...

	age	attrition	business_travel	daily_rate	department	distance_from_home	education	education_field	employee_number	environment_satisfaction	gender	hourly_rate	job_involvement	job_level	job_role	job_satisfaction	marital_status	monthly_income	monthly_rate	num_com
0	51	No	Unknown	684.0	Research & Development	6	3	Unknown	162.0	1.0	Male	51.0	3	5	Research Director	3	Unknown	19537.0	6462	
1	52	No	Unknown	699.0	Unknown	1	4	Life Sciences	259.0	3.0	Male	65.0	2	5	Manager	3	Unknown	19999.0	5678	
2	42	No	travel_rarely	532.0	Research & Development	4	2	Technical Degree	319.0	3.0	Male	58.0	3	5	Manager	4	Married	19232.0	4933	
3	47	No	travel_rarely	359.0	Research & Development	2	4	Medical	Unknown	1.0	Female	82.0	3	4	Research Director	3	Married	17169.0	26703	
4	46	No	Unknown	1319.0	Sales	3	3	Technical Degree	Unknown	1.0	Female	45.0	4	4	Sales Executive	1	Divorced	5820.5	7739	
5	48	No	Unknown	117.0	Research & Development	22	3	Medical	1900.0	4.0	Female	58.0	3	4	Manager	4	Unknown	17174.0	2437	
6	59	No	Unknown	1435.0	Sales	25	3	Life Sciences	81.0	1.0	Female	99.0	3	3	Sales Executive	1	Unknown	5820.5	2354	
7	42	No	travel_rarely	635.0	Sales	1	1	Unknown	387.0	2.0	Male	99.0	3	2	Sales Executive	3	Married	5820.5	24532	
8	41	No	Unknown	1276.0	Unknown	2	5	Unknown	Unknown	2.0	Female	91.0	3	4	Manager	1	Married	16595.0	5626	
9	41	No	travel_frequently	840.0	Research & Development	9	3	Unknown	999.0	1.0	Male	64.0	3	5	Research Director	3	Unknown	16704.0	3735	

...

Last 10 rows of the DataFrame:

...

	age	attrition	business_travel	daily_rate	department	distance_from_home	education	education_field	employee_number	environment_satisfaction	gender	hourly_rate	job_involvement	job_level	job_role	job_satisfaction	marital_status	monthly_income	monthly_rate	n
1604	41	Yes	Unknown	1085.0	Research & Development	2	4	Life Sciences	927.0	2.0	Female	66.028105	1	1	Laboratory Technician	4	Divorced	3034.0	17725	
1605	31	No	travel_rarely	154.0	Sales	7	4	Unknown	941.0	2.0	Male	66.028105	2	1	Sales Representative	3	Unknown	2329.0	11737	
1606	26	No	travel_frequently	1283.0	Sales	1	3	Medical	956.0	3.0	Male	52.000000	2	2	Sales Executive	1	Unknown	4294.0	11148	
1607	31	No	Unknown	616.0	Research & Development	12	3	Medical	Unknown	4.0	Female	41.000000	3	2	Healthcare Representative	4	Unknown	6755.0	17369	
1608	32	No	Unknown	498.0	Research & Development	3	4	Medical	966.0	3.0	Female	93.000000	3	2	Manufacturing Director	1	Married	6725.0	13554	
1609	36	Yes	travel_rarely	530.0	Sales	3	1	Life Sciences	967.0	3.0	Male	51.000000	2	3	Sales Executive	4	Married	10325.0	5518	

Extracción, Transformación y Carga de los datos (ETL)

Automatización del Flujo de Datos

Extracción y Transformación de datos
para asegurar integridad y calidad

Carga de Datos en Base de Datos

Garantía de eficiencia y consistencia en la
actualización de la base de datos.



File

Edit

Selection

View

Go

Run

Terminal

Help

←

→

Proyecto3_AnalisisRRHH

📄

🔍

🔗

📁

🖨

🧪

🔄

👤

⚙

EXPLORER

OPEN EDITORS 1 unsaved

GROUP 1

main.py proyecto-da_promoi_m... 1

GROUP 2

Interactive-1

PROYECTO3_ANALISISRRHH

proyecto-da_promoi_modulo2_te...

data

images

jupyter_notebooks

src

.gitignore

conclusions_exploratory_data_analysis...

DB_initial_structure.mwb

main.py 1

README.md

.gitignore

Data_elena.ipynb

DB_initial_structure.mwb

EDA_elena.ipynb

hr_data_transformed_elena.csv

HR_elena.csv

script_db.sql

OUTLINE

TIMELINE

main.py 1

proyecto-da_promoi_modulo2_team1 > main.py > ...

1 # Imports

2 Run Cell | Run Below | Debug Cell

2 ###

3 import pandas as pd

4 from src.data_transformer import DataTransformer as data

5 from src.database_creation import CreateDatabase as db

6 from src import queries as query

7 from datetime import datetime

8

9

10 # FIRST STEP: EXTRACT DATA

11

12 df = pd.read_csv("data/HR RAW DATA.csv", index_col=0)

13 df.head()

14 hr_data = pd.read_csv("data/hr_data_transformed_2024-08-02_16...

15

16

17

18 # SECOND STEP: TRANSFORMATION

19 abc_data = data(df)

20

21 abc_data.rename_columns() # rename columns

22

23 abc_data.convert_age_to_numbers() # convert number written in

24

25 abc_data.convert_to_numeric("age","integer") # convert colum

26

27 abc_data.fix_negative_distances() # fix negative data into a

28

29 abc_data.correct_env_satisfaction_values() # correct data in

30

31 abc_data.replace_gender_values() # replace wrong values

32

33 abc_data.correct_hourly_rate() # replace wrong values

34

35 abc_data.transform_to_float("monthly_income") # transform to

36

Interactive-1

Interrupt | Clear All | Restart | Variables | Save | Python 3.11.undefined

import pandas as pd ...

Press Enter Connecting to kernel: Python 3.11.9: Activating Python Environment 'P...

main*

0 4

0

Live Share

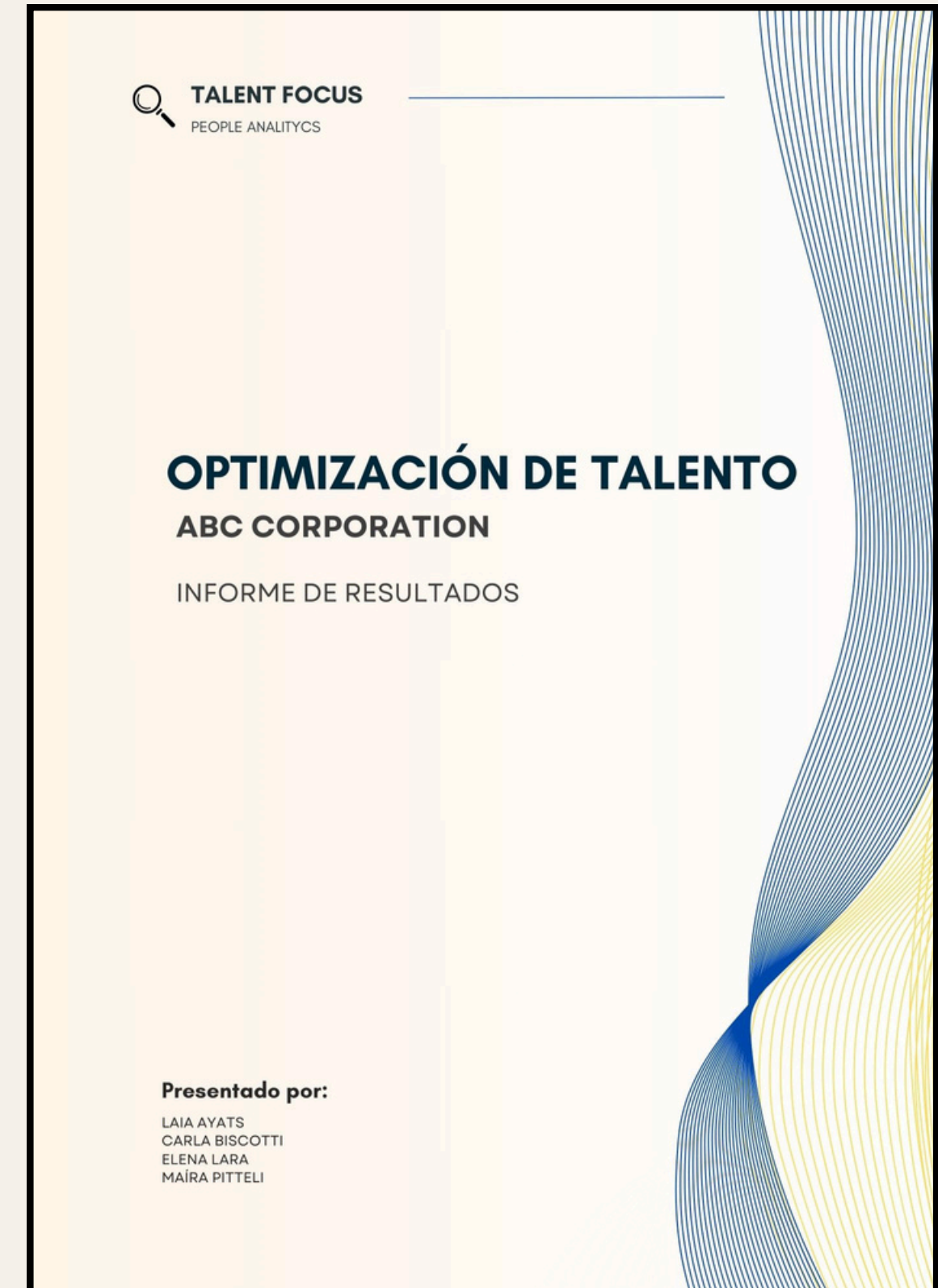
Ln 101, Col 1 Spaces: 4 UTF-8 LF Python 3.12.5 64-bit (Microsoft Store) Go Live

Atletismo Resultado final

Búsqueda

14:51 09/08/2024

Análisis A/B, informe de resultados y sugerencias



Próximos pasos

- Implementar las mejoras sugeridas en en el proceso actual de recolección de datos
- Profundizar análisis e interpretación de los resultados A/B testing considerando otras variables, como “**salaries**” y “**age**”
- Explorar los datos de la **Plataforma de Selección Inteligente** recientemente implementada

Muchas gracias!



APPLAUSE