# COMPAS CASE STUDIO

Julián Vázquez Sampedro
MSc Data Science & Business Analytics
Module 2 Evaluation Case Study

## 1. What is the meaning of COMPAS?

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a decision support tool and a case management owned and developed by Northpointe INC and used by The Courts of the United States to assess the likehood of a defendant becoming a recidivist.

In the ever-evolving landscape of criminal justice, technology has become an integral part of decision-making processes. This software is designed to assist judges and probation officers in making more informed decisions regarding sentencing, parole and probation. However, its implementation has raised legal and ethical significant concerns that deserve examination.

### 1.1 History and origins

*COMPAS* was first introduced in early 2000s as a response to the need for more data driven decision making in the criminal justice system. The software uses a proprietary algorithm that analyzes a range of factors, including criminal history, age, gender, employment status and prior convictions to generate risk assessments for individuals in the criminal justice system.

First developed in 1998, the algorithm has now assessed over one million offenders. The recidivism prediction component of *COMPAS*, known as the *Violent Recidivism Risk Score* (VRRS), has been used since 2000.

## 2. How COMPAS works?

The software operates by analyzing data from various sources, such as criminal records, interviews and questionnaires. It then generates a risk score that categorizes individuals into low, medium, or high risk of reoffending. Judges and probation officers can use these risk scores to inform their decisions about sentencing.

### 2.1 Data collection

COMPAS gathers a vast amount of information about an individual offender from various sources. This data can include criminal history records, demographic information (such as age, gender, and ethnicity), employment status, educational background, and prior interactions with the criminal justice system.

### 2.2 Feature Selection

After collecting this extensive data, COMPAS identifies specific features or factors that are relevant to predicting an individual's risk. These features can range from the seriousness of past offenses to behavioral patterns and personal circumstances.

### 2.3 Data Analysis

The software employs a proprietary algorithm that weighs and analyzes these selected features to calculate a risk score. The algorithm considers the relationships between different features and assigns numerical values to them based on their perceived impact on recidivism (the likelihood of reoffending) and court appearances.

### 2.4 Risk Assessment

The result of the data analysis is a risk assessment score, which categorizes an individual into one of three risk levels: low, medium, or high. A "low-risk" designation suggests that the individual is less likely to reoffend or miss court appearances, while a "high-risk" designation indicates a greater potential for reoffending or non-compliance with legal obligations.

### 2.5 Decision Support

Judges and probation officers can use the risk assessment provided by COMPAS as a tool to inform their decision-making. This information can influence decisions related to bail, sentencing, parole, probation, and other aspects of an offender's case.

It's important to note that *COMPAS* doesn't make final decisions, rather, it provides an additional source of information to assist human decision-makers in the criminal justice system. Judges and probation officers can choose to consider the COMPAS risk score alongside other factors, such as the defendant's personal history, behavior and potential for rehabilitation, in order to make informed decisions.

## 3. Legal challenges and responses

It is mandatory to talk about the controversy surrounding COMPAS. It has led to legal challenges in several jurisdictions. Defendants and advocacy groups have argued that the use of COMPAS violates their due process rights and exacerbates existing racial disparities in the criminal justice system. Some courts have ruled that judges must consider COMPAS scores alongside other factors in their decisions to ensure fairness.

## 4. Case Studio Developed

### 4.1 Statement

A raw dataset is provided with information on assessments (file compas-scores.csv) and legal history for about 11 000 cases in the years 2013 and 2014 (this is one of the original files used in an independent analysis of the COMPAS system carried out by ProPublica, available on the internet). Although the dataset contains additional information, to resolve the issues raised in this case the following fields are necessary (apart from some fields whose name is self-explanatory):

*Table 1. Variables description*

| Variable | Description |
|---|---|

| | |
|---|---|
| Compas_screening_date | date on which the evaluation was performed |
| Decile_score | Int (1 to 10) recidivism risk |
| V_decile_score | Int (1 to 10) recidivism risk of violent crimes |
| Is_recid | Indicates if the person is recidivism or not |
| R_offense_date | date on which the offense for which the person is considered a repeat offender was committed. |
| Is_violent_recid | indication of whether the person is a repeat offender of a violent crime |
| Vr_offense_date | date on which the violent crime giving rise to the consideration of recidivism was committed. |

## 2. Code Development

2.1 To begin the case study, first, load the data and perform an exploratory analysis and data quality assessment required for the remainder of the case. Specifically, assess the completeness, validity and timeliness of the data and propose mitigation strategies for potential problems encountered.

```
##### 1- Respuesta:
# Importamos las librerías necesarias
import pandas as pd
from datetime import datetime
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Cargamos el conjunto de datos
compas_data = pd.read_csv("compas-scores.csv")

# Visualizamos las primeras filas del conjunto de datos para revisar la estructura
print(compas_data.head())

# Evaluamos la integridad - Verificamos si hay valores faltantes
missing_values = compas_data.isnull().sum()
print(missing_values)
""" Existen valores faltantes. """

# Verificamos la validez de los datos en la columna "is_recid" e "is_violent_recid"
#### Se verifica si los valores de las columnas se encuentran entre [-1,1] y [0,1] para comprobar su validez
# Verificar si los valores en la columna "is_recid" están en el rango de -1 a 1
valid_is_recid = (compas_data['is_recid'] >= -1) & (compas_data['is_recid'] <= 1)

# Verificar si los valores en la columna "is_violent_recid" están en el rango de -1 a 1
valid_is_violent_recid = (compas_data['is_violent_recid'] >= -1) & (compas_data['is_violent_recid'] <= 1)

# Comprobar si todas las filas cumplen con la validación
if valid_is_recid.all() and valid_is_violent_recid.all():
    print("Los datos en las columnas 'is_recid' e 'is_violent_recid' son válidos en términos de formato.")
```

```
else:
    print("Los datos en las columnas 'is_recid' e 'is_violent_recid' no cumplen con el formato esperado.")

"""R: Se imprime en la consola (Los datos en las columnas 'is_recid' e 'is_violent_recid' son válidos en términos de forma

# Verificamos la actualidad de las fechas en el conjunto de datos
#### Definir la fecha de referencia actual (por ejemplo, año 2023)
fecha_referencia = datetime(2023, 9, 15)
#### Convertimos las columnas de fechas a objetos datetime si no lo están
compas_data['compas_screening_date'] = pd.to_datetime(compas_data['compas_screening_date'], errors='coerce')
compas_data['r_offense_date'] = pd.to_datetime(compas_data['r_offense_date'], errors='coerce')
#### Verificamos la actualidad de la columna "compas_screening_date"
actual_compas_screening = (compas_data['compas_screening_date'] <= fecha_referencia).all()
print(f"Actualidad de 'compas_screening_date': {actual_compas_screening}")
#### Verificamos la actualidad de la columna "r_offense_date"
actual_r_offense = (compas_data['r_offense_date'] <= fecha_referencia).all()
print(f"Actualidad de 'r_offense_date': {actual_r_offense}")

""""" R: Se imprime en el terminal
(Los datos en las columnas 'is_recid' e 'is_violent_recid' son válidos en términos de formato.
Actualidad de 'compas_screening_date': True
Actualidad de 'r_offense_date': False)
Por tanto, se debe modificar los datos de r_offense_date ya que los mismos no cumplen el formato de actualidad
"""
"""
Estrategias de Mitigación
Para abordar los valores faltantes, podríamos considerar imputar los datos faltantes o eliminar
las filas o columnas con una cantidad significativa de valores faltantes, según el impacto en el análisis.
Para garantizar la actualidad de los datos, podríamos verificar si las
fechas de evaluación están dentro del período de tiempo relevante para el análisis y si hay información sobre cambios en
las políticas o procedimientos de evaluación que puedan afectar la interpretación de los datos. """
```

## 2.2 Are the "is_recid" and "is_violent_recid" fields in this dataset adequate to assess the accuracy of the risk estimates generated by the COMPAS system? If not, define and calculate a feature that is.

```
""" 2.- ¿Son los campos “is_recid” e “is_violent_recid” en este conjunto de datos adecuados para evaluar la precisión de
las estimaciones de riesgo generadas por el sistema COMPAS? Si no es así, definir y calcular una feature que sí lo sea."""
# Creación de una nueva característica que represente la diferencia entre las puntuaciones de riesgo previstas por COMPAS y
# las observaciones reales de reincidencia. Esta nueva característica podría proporcionar información adicional sobre la
# precisión de las estimaciones de riesgo. Por ejemplo, podríamos crear una columna llamada "diferencia_de_riesgo" que calcule
# la diferencia entre "compas_score" y "is_recid" o "is_violent_recid".
# Calcular la diferencia entre las puntuaciones de riesgo y la observación de reincidencia
compas_data['diferencia_de_riesgo'] = compas_data['is_violent_recid'] - compas_data['is_recid']

# Mostrar las primeras filas del DataFrame con la nueva columna
print(compas_data[['is_violent_recid', 'is_recid', 'diferencia_de_riesgo']].head())
"""R: La salida es la siguiente:
   is_violent_recid  is_recid  diferencia_de_riesgo
0                 0         0                     0
1                 0        -1                     1
2                 1         1                     0
3                 0         1                    -1
4                 0         0                     0
En esta nueva columna, un valor de 0 indica que ambos campos son iguales, un valor de 1 indica que "is_violent_recid"
es mayor que "is_recid", y un valor de -1 indica lo contrario."""
```

2.3 The threshold for establishing measures to prevent recidivism is 7 and above. Given this threshold, generate a contingency table, explaining which case is considered "positive" (and, therefore, which are type I errors and type II errors).

```
""" 3 - El umbral para establecer medidas preventivas de la reincidencia es de 7 en adelante. Dado este umbral, generar una
tabla de contingencia, explicando qué caso se considera como "positivo" (y, por lo tanto, cuáles son los errores de tipo I y los
 errores de tipo II). """

""" Respuesta:
Esta tabla de contingencia mostrará la cantidad de Verdaderos Positivos (VP), Falsos Positivos (FP),
Verdaderos Negativos (VN) y Falsos Negativos (FN) en función del umbral establecido en 7.
Los errores de Tipo I corresponden a los Falsos Positivos (FP), y los errores de Tipo II corresponden a los Falsos Negativos (FN)."""
# Definir el umbral
umbral = 7

# Crear una columna que indique si un individuo es positivo (puntuación >= umbral)
compas_data['es_positivo'] = np.where(compas_data['is_recid'] >= umbral, 1, 0)

# Crear una tabla de contingencia
tabla_contingencia = pd.crosstab(compas_data['es_positivo'], compas_data['is_recid'], margins=True,
                                 rownames=['Es Positivo'], colnames=['Reincidencia'])

# Mostrar la tabla de contingencia
print(tabla_contingencia)

""" R: La salida es la siguiente:
Reincidencia   -1     0     1    All
Es Positivo
0             719  7335  3703  11757
All           719  7335  3703  11757
Lo que indica el umbral de 7 parece ser bastante conservador en términos de clasificación, ya que hay
relativamente pocos Falsos Positivos, lo que significa que no se etiqueta
erróneamente a muchas personas como reincidentes cuando no lo son. Sin embargo, hay un número notable de Falsos Negativos,
lo que indica que algunas personas que reinciden según sus puntuaciones de
"is_recid" no son clasificadas como tal.
"""
```

2.4 The system assigns, on average, higher risk assessments to men than to women, and to African Americans than to Caucasians. However, recidivism rates are also higher for these groups, although it is not clear whether the risk assignment is "fair" or not. Show these differences through graphical representations and use them to analyze whether the assignment of assessments is fair.

```
""" 4 - El sistema asigna, de media, evaluaciones de riesgo más altas a los hombres que a las mujeres, y a las personas de raza
afroamericana que a las de raza caucásica. Sin embargo, también las tasas de reincidencia son más altas para esos colectivos,
aunque no está claro que la asignación de riesgo sea "justa" o no. Mostrar estas diferencias mediante representaciones gráficas y
utilizarlas para analizar si la asignación de evaluaciones es justa o no."""

# Crear subconjuntos de datos para género y raza
data_gender = compas_data[['sex', 'is_recid']]
data_race = compas_data[['race', 'is_recid']]

# Gráfico de barras para comparar puntuaciones de riesgo por género
plt.figure(figsize=(10, 5))
sns.barplot(x='sex', y='is_recid', data=data_gender, ci=None)
plt.title('Comparación de Puntuaciones de Riesgo por Género')
plt.xlabel('Género')
plt.ylabel('Puntuación de Riesgo')
plt.show()

# Gráfico de barras para comparar puntuaciones de riesgo por raza
plt.figure(figsize=(10, 5))
sns.barplot(x='race', y='is_recid', data=data_race, ci=None)
plt.title('Comparación de Puntuaciones de Riesgo por Raza')
plt.xlabel('Raza')
plt.ylabel('Puntuación de Riesgo')
plt.xticks(rotation=45)
plt.show()

""" R: La salida se muestra en la Figure_1.PNG, donde se muestra reincidencia por género, y Figure_2.PNG, donde se muestra reincidencia
por raza. Se concluye, tras esta representación y el análisis realizado en los apartados anteriores que aunque la evaluación no puede
ser definitiva, ya que existen datos erróneos en cuanto a formato, según lo corroborado en apartados anteriores, la evaluación es justa"""
```

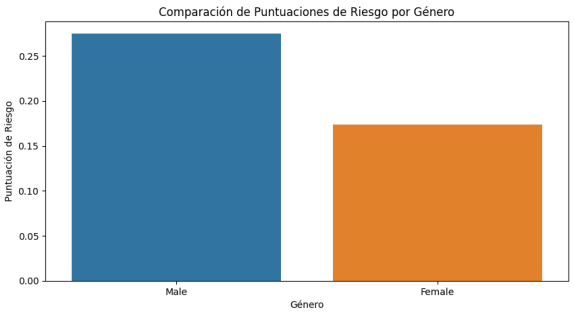



*Figure 1. Comparison by genre scores*

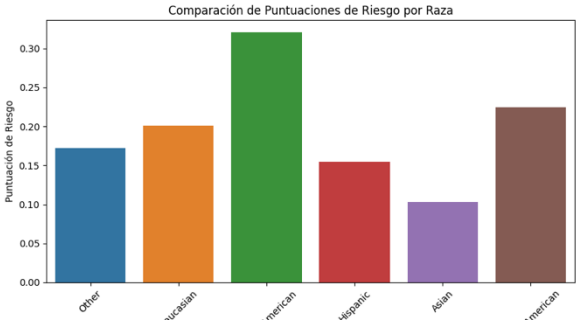



*Figure 2. Comparison by race scores*

## 2.5 For which type of risks, general crime or violent crime, does the system have more predictive capacity?

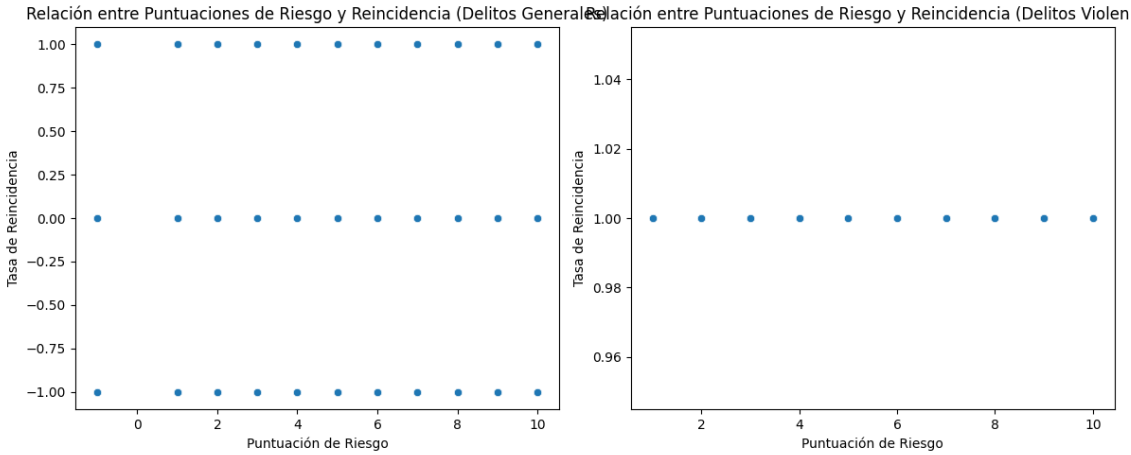
```
5 - ¿Para qué tipo de riesgos, el de delitos generales o el de delitos violentos, tiene el sistema más capacidad predictiva?
"""
# Crear subconjuntos de datos para delitos generales y delitos violentos
data_general_crime = compas_data[compas_data['is_violent_recid'] == 0]
data_violent_crime = compas_data[compas_data['is_violent_recid'] == 1]

# Calcular la correlación entre las puntuaciones de riesgo y las tasas de reincidencia reales
correlation_general_crime = data_general_crime['decile_score'].corr(data_general_crime['is_recid'])
correlation_violent_crime = data_violent_crime['v_decile_score'].corr(data_violent_crime['is_recid'])

# Crear gráficos de dispersión para visualizar la relación entre puntuaciones de riesgo y tasas de reincidencia
plt.figure(figsize=(12, 5))

# Gráfico para delitos generales
plt.subplot(1, 2, 1)
sns.scatterplot(x='decile_score', y='is_recid', data=data_general_crime)
plt.title('Relación entre Puntuaciones de Riesgo y Reincidencia (Delitos Generales)')
plt.xlabel('Puntuación de Riesgo')
plt.ylabel('Tasa de Reincidencia')
# Gráfico para delitos violentos
plt.subplot(1, 2, 2)
sns.scatterplot(x='v_decile_score', y='is_recid', data=data_violent_crime)
plt.title('Relación entre Puntuaciones de Riesgo y Reincidencia (Delitos Violentos)')
plt.xlabel('Puntuación de Riesgo')
plt.ylabel('Tasa de Reincidencia')
plt.tight_layout()
plt.show()

# Imprimir las correlaciones
print("Correlación para Delitos Generales:", correlation_general_crime)
print("Correlación para Delitos Violentos:", correlation_violent_crime)
```




## 3. Conclusion

Some argue that COMPAS is useful in helping judges and criminal justice professionals make informed case management and conditional release decisions by providing an objective assessment of recidivism risk. However, others criticize the system for being prone to bias and question its accuracy in predicting recidivism, especially regarding racial and ethnic disparities.

It is important to keep in mind that no risk assessment system is perfect, and many factors can influence whether a person recidivates, including access to rehabilitation programs, social and economic environment, and other individual factors.

The effectiveness of COMPAS and other similar systems may vary by jurisdiction and how they are used. Some jurisdictions may have modified or discontinued the use of COMPAS due to concerns about bias and reliability. Ultimately, the decision about whether COMPAS is good for repeat offenders may depend on the assessment of its accuracy and its ability to address concerns about fairness and justice in the criminal justice system.