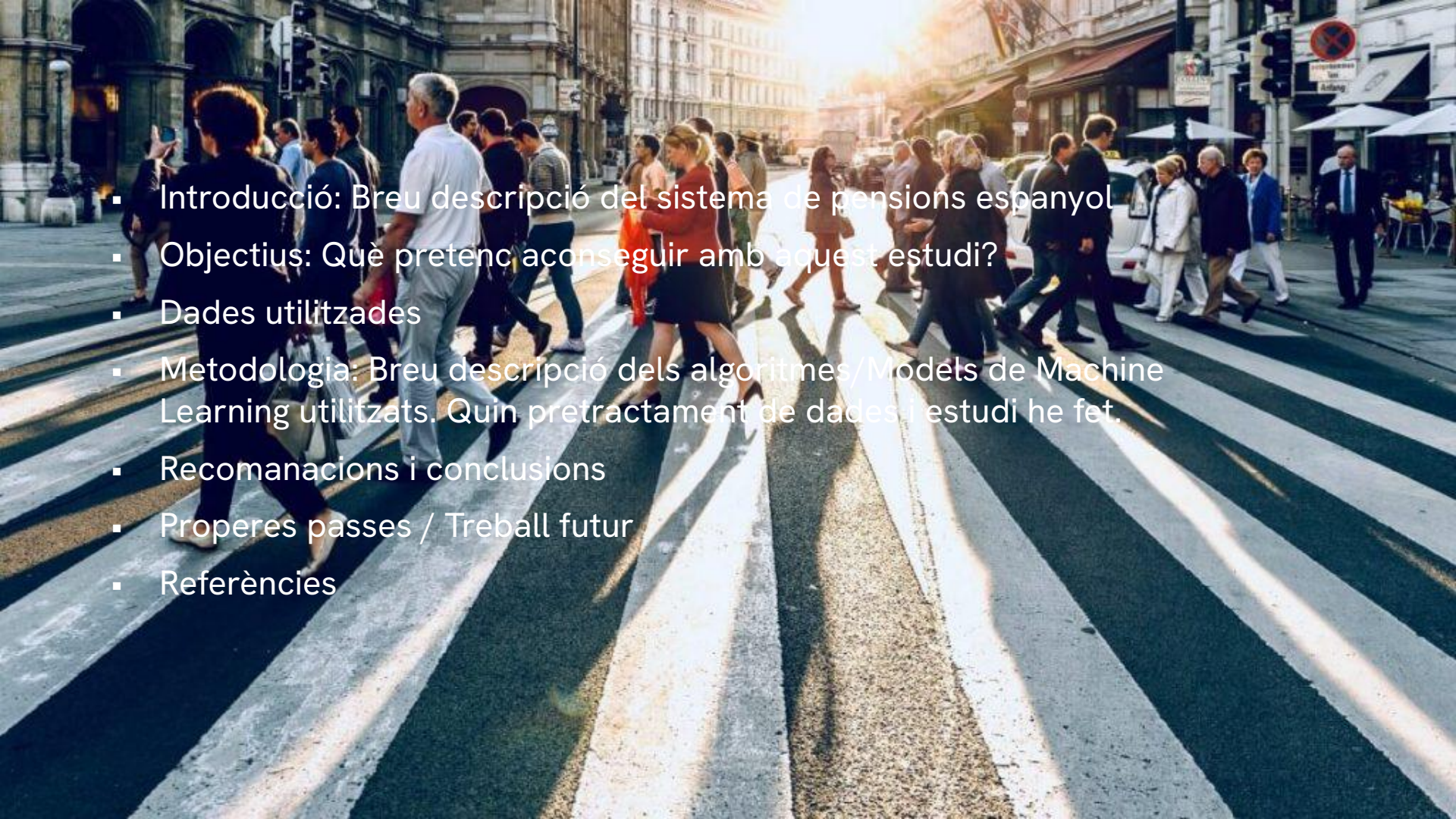




Machine Learning en la sostenibilitat del Sistema de pensions public



- 
- Introducció: Breu descripció del sistema de pensions espanyol
 - Objectius: Què pretenc aconseguir amb aquest estudi?
 - Dades utilitzades
 - Metodologia: Breu descripció dels algoritmes/Models de Machine Learning utilitzats. Quin pretractament de dades i estudi he fet.
 - Recomanacions i conclusions
 - Properes passes / Treball futur
 - Referències

Introducció: Breu descripció del sistema de pensions espanyol

El sistema de pensions espanyol, se l'anomena (**sistema de reparto o de "solidaridad intergeneracional"**) que no existeix un FONS o Guardiola de pensions.

-La actual generació cotitzant, es financida per la generació que li segueix i que actualment està contribuïnt/treballant.

La despesa en impostos per treballador es de les més altes de la OCDE i donen poc marge per incrementar aquest % i és per això que els diferents partits polítics segueixen incrementant el deute, ja que no son mesures populars a aplicar entre la població ni a les empreses...

Els reptes actuals del sistema de pensions espanyol inclouen l'envelliment de la població, el desequilibri entre les ingressos de cotitzants i pensions i l'augment de l'esperança de vida, que posen en perill la viabilitat del sistema.

Objectius: Què pretenem aconseguir amb aquest estudi?

Predir la sostenibilitat del sistema a llarg termini.

Identificar els factors clau que afecten el sistema de pensions.

Proporcionar recomanacions per millorar la gestió del sistema de pensions.

Deuda de la Seguridad Social

Deuda de la Seguridad en millones de euros



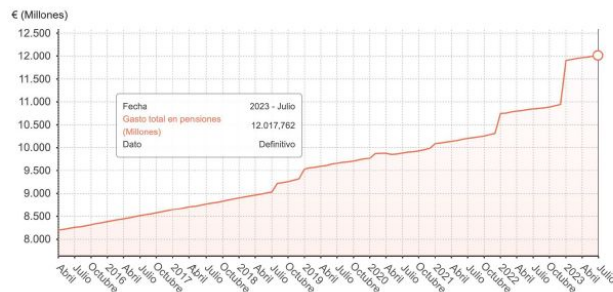
Gráfico: El Confidencial · Fuente: Banco de España · [Descargar los datos](#) · Creado con [Datawrapper](#)

ep data

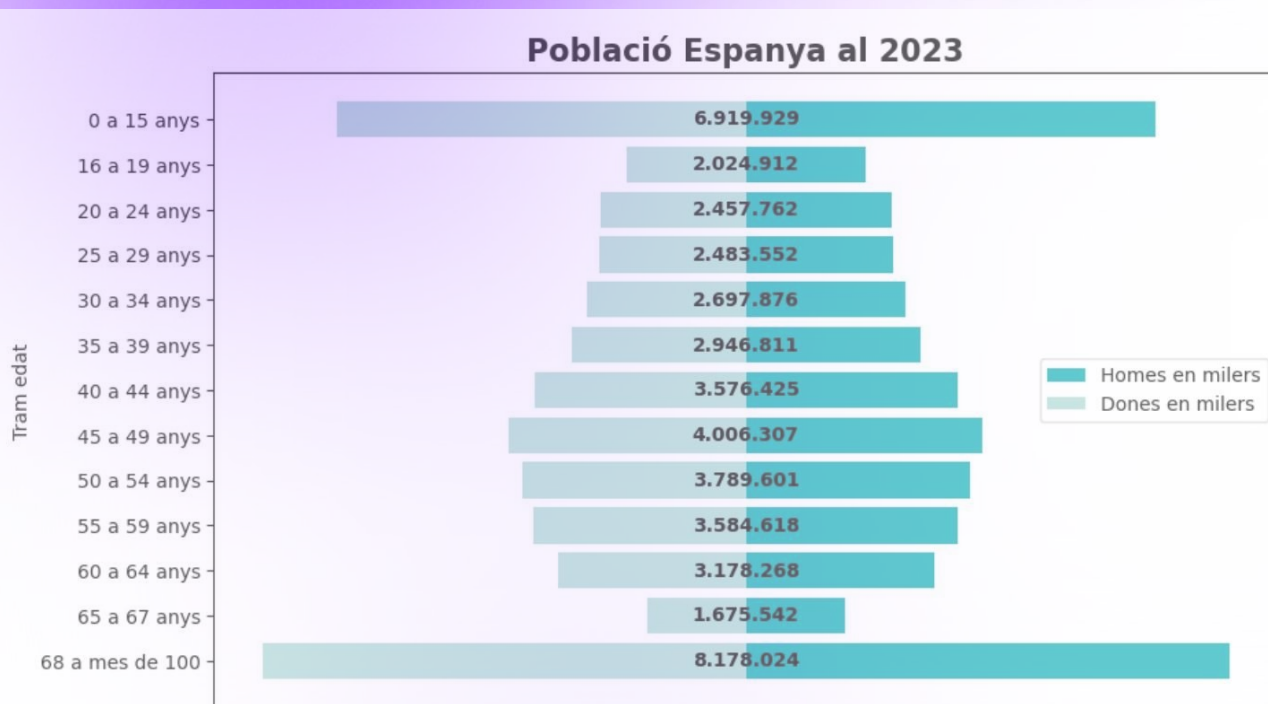
Iniciar Sesión Crear

Evolución del gasto en pensiones en España

Gasto total en pensiones en España



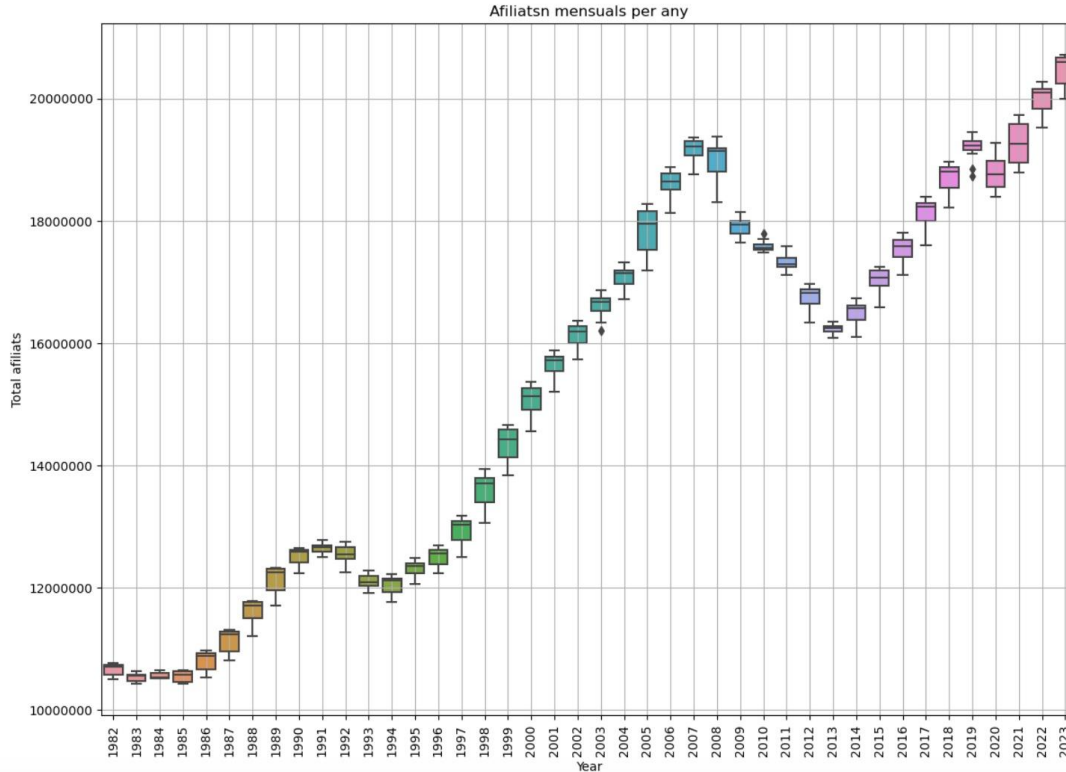
Fonts utilitzades



- Les dades utilitzades per a l'estudi provenen de l'Institut Nacional d'Estadística (INE), l'Institut Nacional d'Estadística (INE), l'Organització de les Nacions Unides (ONU), la Seguretat Social d'Espanya i el Ministeri de Treball i el Ministeri d'Hisenda
- La qualitat de les dades és alta, ja que aquests organismes garanteixen l'exactitud i la fiabilitat
- El període de temps cobert és de 105 anys, des del 1995 fins al 2100
- Les característiques utilitzades inclouen l'edat, el gènere, el tipus de pensió i la quantitat de pensió

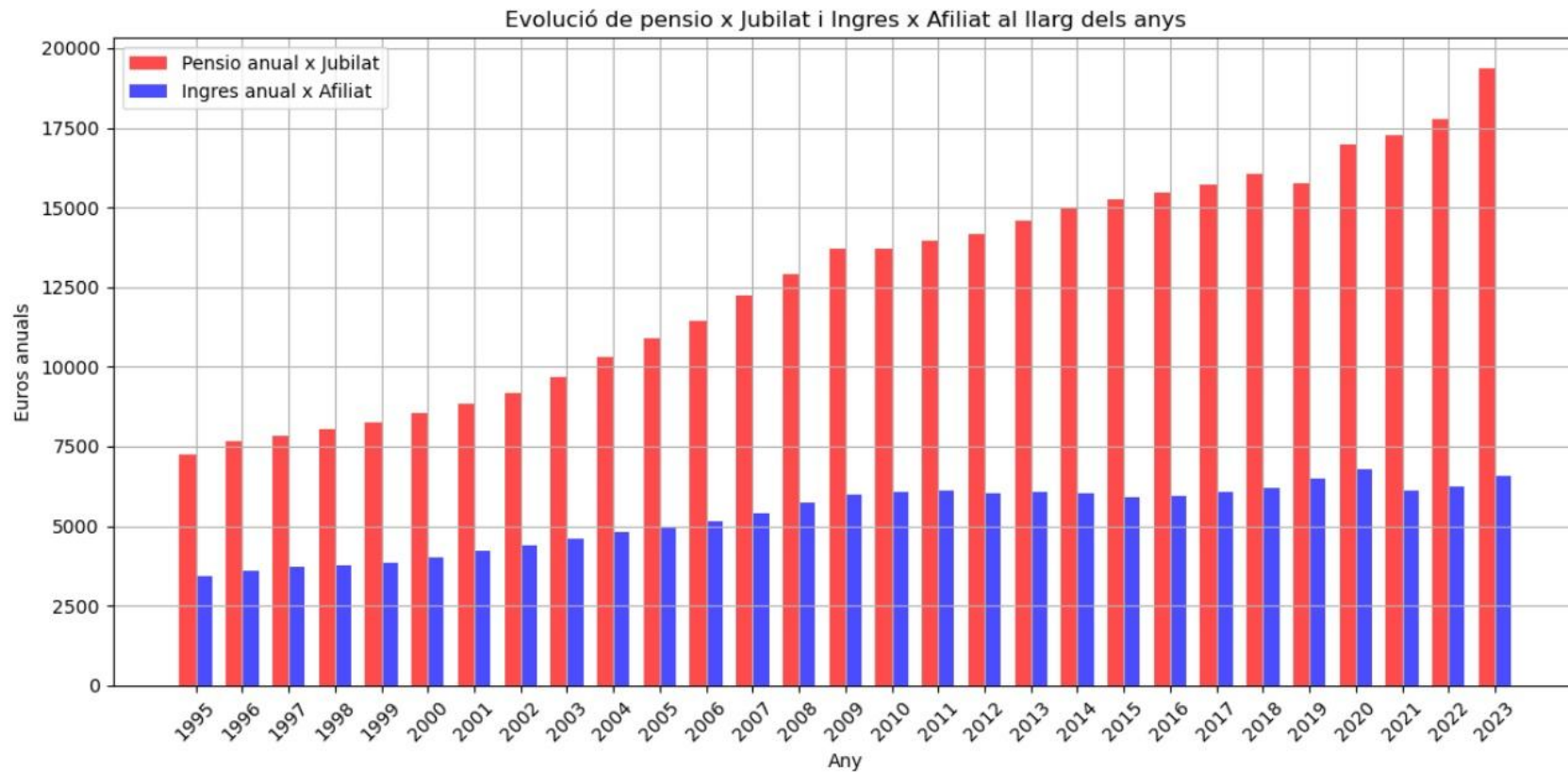
Magnituds : Envelliment de la població

Anàlisi de les dades fins a dia d'avui:



- Hi ha estacionalitat de les dades d'afiliació
- Existeix correlació amb les èpoques de crisi. alta, ja que aquests organismes garanteixen l'exactitud i la fiabilitat

Anàlisi de les dades fins a dia d'avui:



Magnituds : Despesa pensions vs Sanitat i Educació

EscenariA Optimista:

- El Deute no incrementa mes
- La pensió x Jubilats es manté com al 2023 (la més alta de la història)
- Els ingressos per cotitzant augmenten un 2% anual

EscenariB Pessimista:

- El % d'afiliats per població (43,04%) es manté igual que al 2023
- El Ingres x afiliat NO AUMENTA ES QUEDA IGUAL QUE AL 2023
- El deute per càpita (32419€) SEGUEIX INCREMENTANT AMB LA MATEIXA TENDÈNCIA
- El Deute en M€ SEGUEIX INCREMENTANT AMB LA MATEIXA TENDÈNCIA
- El PIB en M€ SEGUEIX INCREMENTANT AMB LA MATEIXA TENDÈNCIA

Metodologia: Breu descripció dels algoritmes/Models de Machine Learning utilitzats.

**Regressió
Lineal**

Algoritme 1

**Random
Forest**

Algoritme 2

**MultiOutput
Regressor**

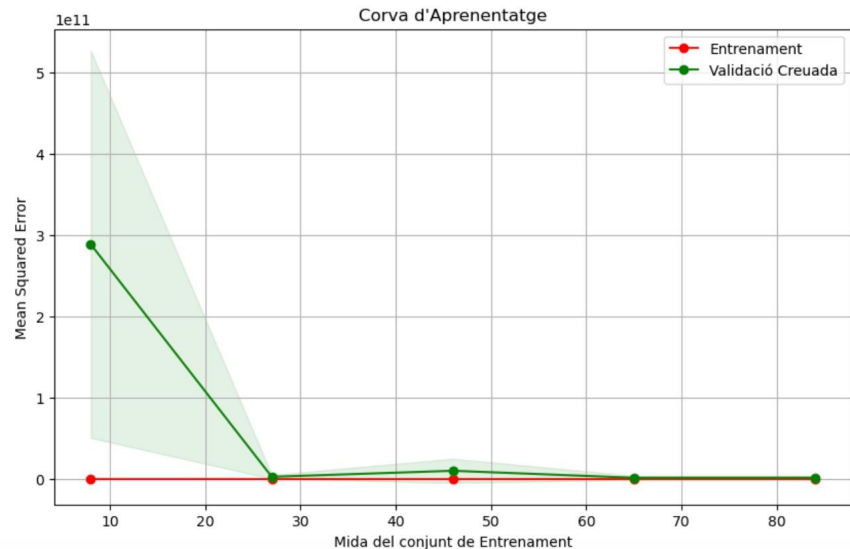
Algoritme 3

Ridge Model

Algoritme 4

Regularització, selecció de característiques, validació creuada

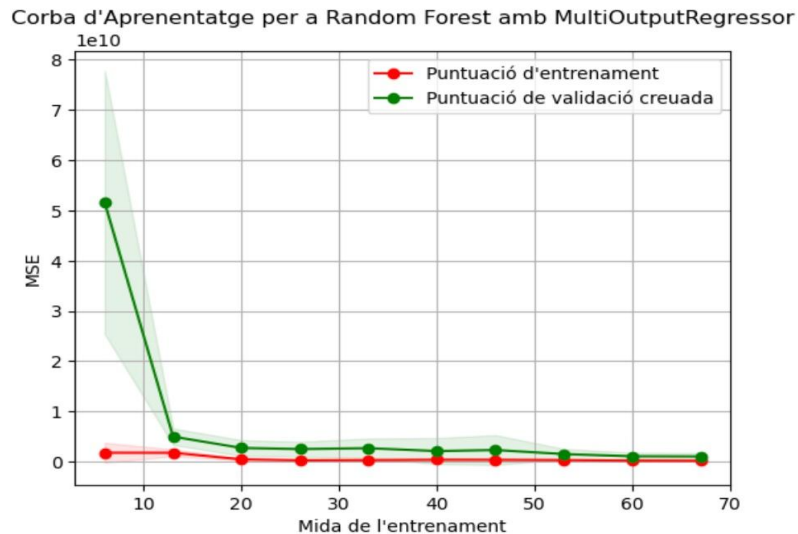
Error de Entrenament: 18137.711353027342
Error de Test: 50996.42417563174



MSE Train (Linear Regression): 15161.476351166159 \pm 5120.817582154896

MSE Test (Linear Regression): 253231.0852343417 \pm 497962.1253850461

Error de Entrenament: 145618968.1997166
Error de Test: 563279870.5131682



MSE Train (MultiOutput): 444016319.19615555 \pm 135304424.54541644

MSE Test (MultiOutput): 1039291759.4155493 \pm 834003775.1658822

Metodologia: Quin pretractament de dades hem fet?

**Boxplots
per veure
outliers**

Algunes variables
tenen outliers

**Subplots per
veure
distribució**

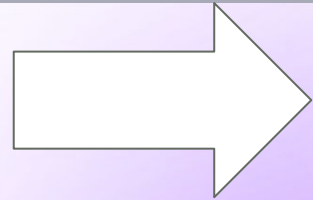
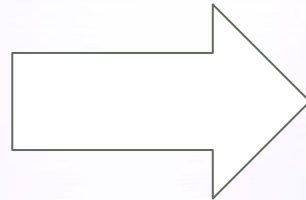
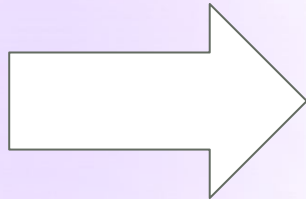
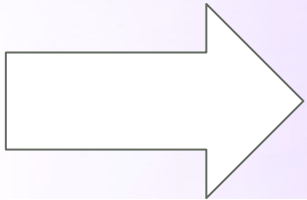
Distribucions no
gaussianas

**Shapiro per
veure
pvalue
(distribució)**

Totes tenen pvalue < 0.05

**Robust
Scaler**

Escalat



Avaluació Ridge o Lasso

```
In [63]: #Definir un conjunt de valors d'alpha (paràmetre de regularització) per ser probat. Un valor més gr
alphas = np.logspace(-6, 6, 13)
#Fem una validació creuada per trobar el millor alpha:
ridge = Ridge()
parameters = {'alpha': alphas}
ridge_regressor = GridSearchCV(ridge, parameters, scoring='neg_mean_squared_error', cv=5)
ridge_regressor.fit(X_train_scaled, y_train)
print("Millor paràmetre Ridge:", ridge_regressor.best_params_)
print("MSERidge:", -ridge_regressor.best_score_)
```

```
Millor paràmetre Ridge: {'alpha': 1e-05}
MSERidge: 173503.0024432324
```

```
In [64]: lasso = Lasso()
parameters = {'alpha': alphas}
lasso_regressor = GridSearchCV(lasso, parameters, scoring='neg_mean_squared_error', cv=5)
lasso_regressor.fit(X_train_scaled, y_train)
print("Millor paràmetre Lasso:", lasso_regressor.best_params_)
print("MSELasso:", -lasso_regressor.best_score_)
```

```
Millor paràmetre Lasso: {'alpha': 1.0}
MSELasso: 1115728.1438143249
```


Prediccions finals amb els Escenari A

```
# Primer, prenem les dades d'entrenament:
y = combined_data['Despesa pensions en M€']
X = combined_data.drop(columns=['Despesa pensions en M€'])

# Entrenem el model Ridge amb el millor valor d'alpha que hem trobat:
ridge_model = Ridge(alpha=1e-05)
ridge_model.fit(X, y)

# Preparar les dades futures per a la predicció:
columns_for_prediction = X.columns
X_future = EscenariA[columns_for_prediction]

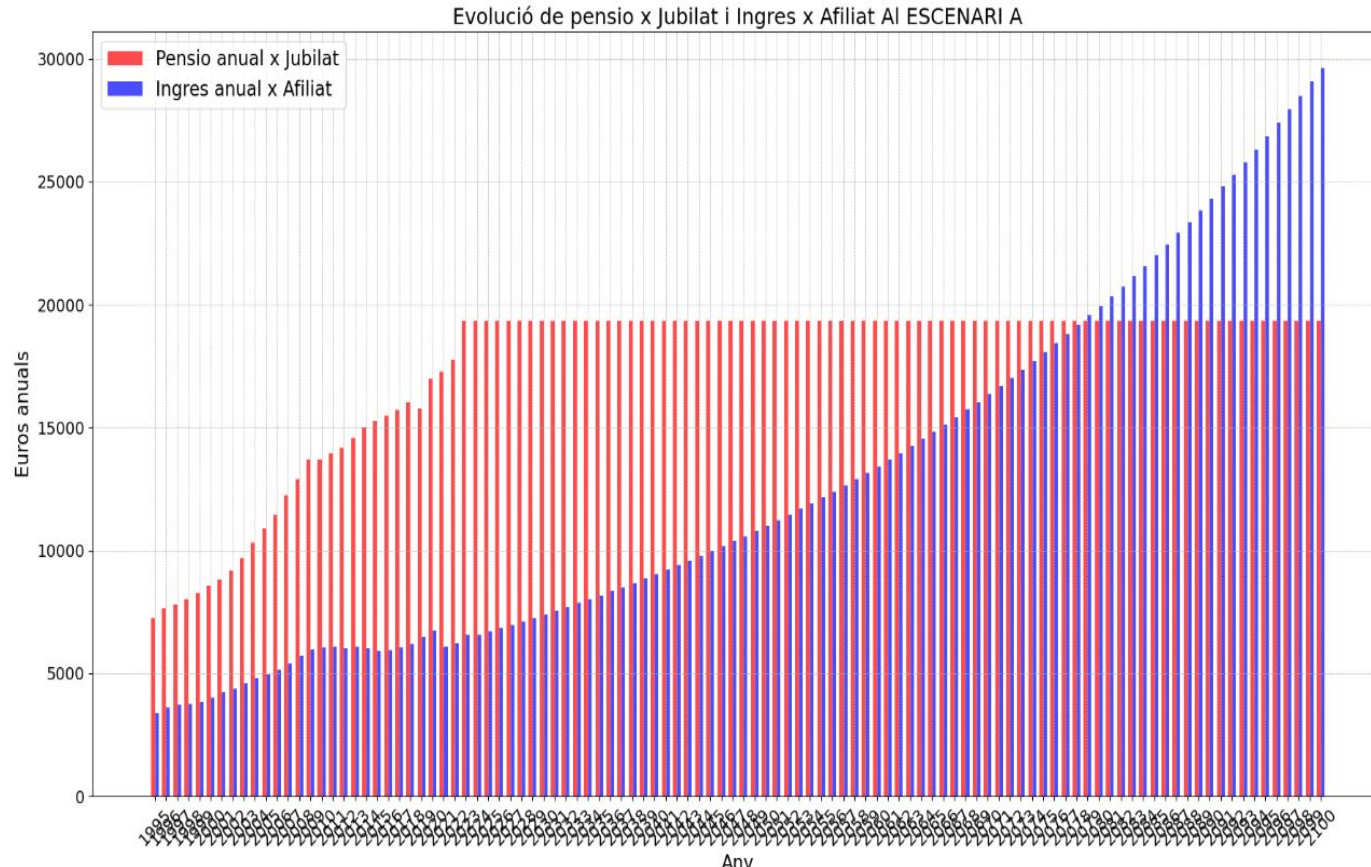
# Fer la predicció amb el model Ridge:
EscenariA['PrediccióRIDGE Despesa pensions'] = ridge_model.predict(X_future)

# Mostrar les primeres files de les prediccions:

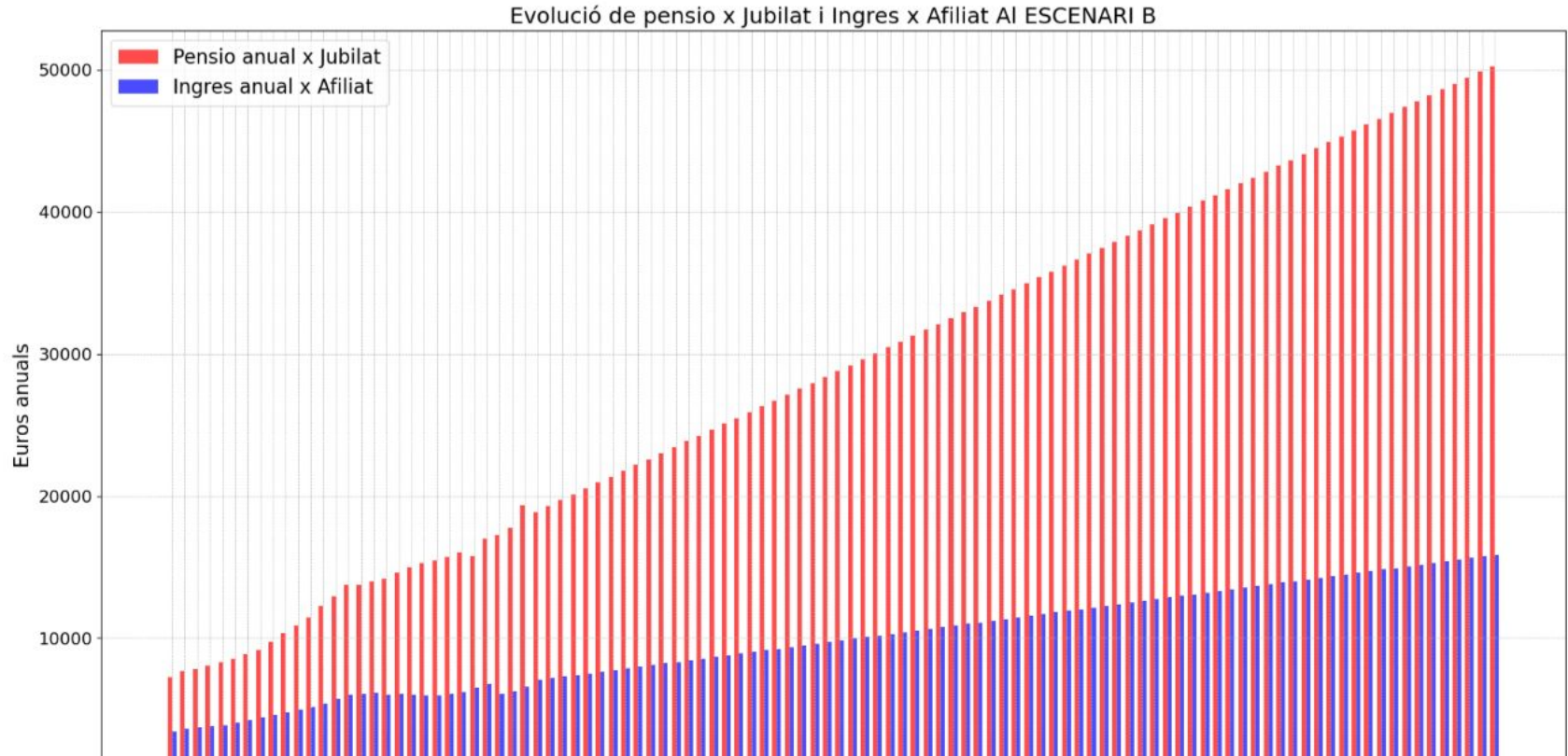
EscenariA.head()
```

| Ingressos cotitzacions en M€ | Fons Reserva en M€ | %PensionsvsPIB | Rati Ingressos vs Despesa | PrediccióLR Despesa pensions | predictedMOR_Despesa pensions en M€ | predictedMOR_Deute en (M€) | PrediccióRIDGE Despesa pensions |
|------------------------------------|--------------------------|----------------|------------------------------------|------------------------------------|--|-------------------------------|---------------------------------------|
| 134446.11 | -392091.43 | 14.17 | 0.69 | 195079.41 | 202714.68 | 1531218.45 | 195021.92 |
| 136980.92 | -454737.24 | 14.52 | 0.69 | 199650.82 | 205389.08 | 1530872.52 | 199616.97 |
| 139545.68 | -519453.12 | 14.88 | 0.68 | 204279.94 | 206724.29 | 1530204.77 | 204273.46 |
| 142142.68 | -586242.34 | 15.24 | 0.68 | 208936.98 | 208099.63 | 1529491.93 | 208889.35 |
| 144772.60 | -655342.68 | 15.62 | 0.68 | 213871.85 | 208388.45 | 1528945.78 | 213857.28 |

Anàlisi de les dades amb els escenari A:



Anàlisi de les dades amb els escenari B:



Prediccions 2 escenaris:

Posa un any a partir del 2024 inclòs: 2037

Comparació de dades entre l'any 2037 i l'any 2023 amb l'EscenariA:

Any: 2037.0 (2023: 2023.0)

Sum_0_to_100+: 46414915.0 (2023: 47519627.0)

Jubilats: 13591955.999999998 (2023: 9853566.0)

% Jubilats: 29.283595585600015 (2023: 20.73578144878957)

% Afiliats: 43.04 (2023: 43.04335072808066)

Afiliats: 19976979.416 (2023: 20454039.714285713)

CotitzantxJubilat: 1.47 (2023: 2.08)

Deute per capita: 32419.0 (2023: 32419.0)

Deute en (M€): 1504725.129385 (2023: 1558149.0)

PIB per capita: 28989.12 (2023: 28989.115183686252)

PIB en M€: 1345527.5407248 (2023: 1377551.9405888072)

Deute(%PIB): 111.8316113079666 (2023: 113.11)

Pensio x Jubilat: 19352.08 (2023: 19352.080251961575)

Despesa pensions en M€: 263032.61986847996 (2023: 190687.0)

Ingres x Afiliat: 8511.931628385983 (2023: 6580.7538207716125)

Ingressos cotitzacions en M€: 170042.68293066614 (2023: 134603.0)

Fons Reserva en M€: -1398429.5798873818 (2023: -331489.0)

%PensionsvsPIB: 19.54866116874808 (2023: 13.842454457179642)

Rati Ingressos vs Despesa: 0.6464699435974515 (2023: 0.7058845123159943)

Posa un any a partir del 2024 inclòs: 2037

Comparació de dades entre l'any 2037 i l'any 2023 amb EscenariB

Any: 2037.0 (2023: 2023.0)

Sum_0_to_100+: 46414915.0 (2023: 47519627.0)

Jubilats: 13591955.999999998 (2023: 9853566.0)

% Jubilats: 29.283595585600015 (2023: 20.73578144878957)

% Afiliats: 43.04 (2023: 43.04335072808066)

Afiliats: 19976979.416 (2023: 20454039.714285713)

Pensio x Jubilat: 24249.70941977715 (2023: 19352.080251961575)

Ingres x Afiliat: 8558.466036990576 (2023: 6580.7538207716125)

Despesa pensions en M€: 329600.98344639654 (2023: 190687.0)

Ingressos cotitzacions en M€: 170972.29985349585 (2023: 134603.0)

CotitzantxJubilat: 1.4697648679851527 (2023: 2.08)

Deute en (M€): 2107943.7793103456 (2023: 1558149.0)

PIB en M€: 1763567.0714005157 (2023: 1377551.9405888072)

Deute per capita: 45415.22438014474 (2023: 32419.0)

PIB per capita: 37995.697533874954 (2023: 28989.115183686252)

Deute(%PIB): 1.195272815814341 (2023: 113.11)

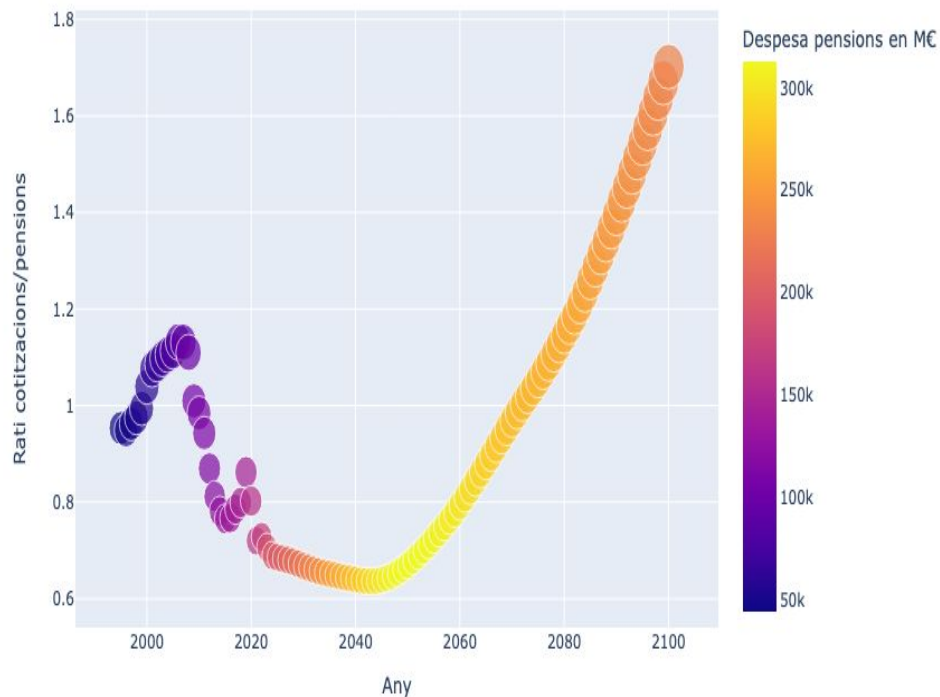
Fons Reserva en M€: -1704481.2438883123 (2023: -331489.0)

%PensionsvsPIB: 18.689449853735805 (2023: 13.842454457179642)

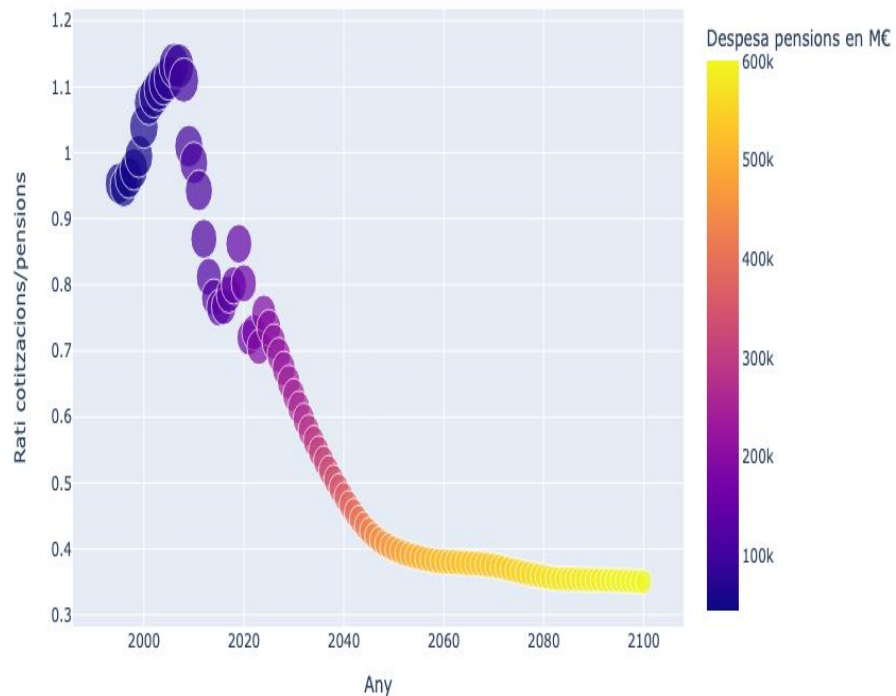
Rati Ingressos vs Despesa: 0.5187250901552644 (2023: 0.7058845123159943)

Prediccions 2 escenaris:

Evolució rati cotitzacions / pensions (1995-2100) Escenari A



Evolució rati cotitzacions / pensions (1995-2100) Escenari B



Temes a tenir en compte, properes passes:

1. La població espanyola començarà a baixar a partir del 2030. Degut a l'envelliment de la població i la mortalitat que aquesta té.
2. S'ha valorat tenir en compte l'efecte de la immigració, però el saldo migratori (immigració-emigració) només ha estat positiu a partir del 2018 , i l'increment en cotitzacions no ha estat significatiu.
3. Hi ha molt biaix a les dades i notícies sobre aquest tema. La majoria d'estudis son subvencionats per ents publicos o bancs amb interessos polítics o comercials.S'utilitza a nivell polític per presumir de que el país va bé, però les dades d'endeutament demostren el contrari.

Quines altres anàlisis consideres que serien útils?

1. Anàlisi de les tendències de jubilació anticipada i el seu efecte en el finançament del sistema.

2. Avaluació de la sostenibilitat financera del país.

3.Tenir una app actualitzada per entrar la data de jubilació i veure la pensioxCapita

Quines dades addicionals serien útils?

1. Dades anteriors a 1995 ens donaran una predicció mes exacte. Els models de machine learning funcionen millor amb grans bases de dades més grans ja que els permeten aprendre patrons més fàcilment. Amb els datasets petits es molt fàcil fer overfitting i tenen més probabilitat d'error.

2. Informació sobre les prestacions de pensions actuals, com la quantitat mitjana de pensió i el nombre de beneficiaris reals

3. Nombre dels cotitzants reals i unics (no multiples entrades al sistema) Afiliats que cotitzen 12 mesos a l'any.

Conclusió

- El nombre d'afiliats no es rellevant.
- La conclusió més clara és que el sistema actual de pensions en ambdós escenaris ****NO es sostenible per un problema de manca d'ingressos de cotitzacions****
- No compta amb una guardiola de pensions, tot i que ****els ingressos només han superat les pensions breument desde l'any que tenim dades comptables del 1995****, No hi ha suficients cotitzants per la població existent. Hi ha molta estacionalitat laboral a Espanya i els que cotitzant són a temps parcial i/o amb bases de cotització molt baixes.
- La població espanyola està envellint i la despesa en pensions està augmentant dramàticament,
- El segment de població que està a punt de jubilar-se (baby boomer i gen X) són els que han contribuït més al sistema amb bases més altes i durant més anys.
- **LES PERSPECTIVES D'OBTENIR UNA PENSIÓ PÚBLICA AL 2037 NO HAN CANVIAT DESDE EL 1991....**



<https://www.linkedin.com/in/angelssastre/>



<https://github.com/angeliufus>

Gràcies!

IT ACADEMY