

# DAT102x: Predicting Mortgage Approvals From Government Data

HOSTED BY MICROSOFT

## Descripción del problema

- Sobre los datos
- Variable objetivo
  - Formato de envío
  - Métrica de rendimiento
- Características
  - Fila de ejemplo
- Referencias

## Sobre los datos

Su objetivo es predecir si una solicitud de hipoteca fue aceptada (lo que significa que el préstamo se originó ([https://en.wikipedia.org/wiki/Loan\\_origination](https://en.wikipedia.org/wiki/Loan_origination)) ) o denegada de acuerdo con el conjunto de datos dado, que está adaptado del Consejo de Examen de Instituciones Financieras Federales (FFIEC) (<https://www.ffiec.gov/hmda/hmdaflat.htm>) .

INICIO

(/COMPETITIONS/14/MORTGAGE-APPROVALS-FROM-GOVERNMENT-DATA/PAGE/43/)

DESCRIPCIÓN DEL PROBLEMA

(/COMPETITIONS/14/MORTGAGE-APPROVALS-FROM-GOVERNMENT-DATA/PAGE/44/)

PUNTAJE DE EXPORTACIÓN PARA EDX

(/COMPETITIONS/14/MORTGAGE-APPROVALS-FROM-GOVERNMENT-DATA/PAGE/45/)

LEADERBOARD

(/COMPETITIONS/14/MORTGAGE-APPROVALS-FROM-

# Variable objetivo

Estamos tratando de predecir la variable `accepted` (una variable binaria) para cada fila del conjunto de datos de prueba.

Tu trabajo es:

1. Entrenar un modelo usando las entradas `train_values.csv` y las etiquetas `train_labels.csv`
2. Predecir el valor de cada fila en la `test_values.csv` que no conoce el valor verdadero de `accepted`.
3. Envíe sus predicciones en un formato que coincida `submission_format.csv` **exactamente**.
4. Sube tus predicciones a esta competencia para obtener un puntaje.
5. Exporte su token de calificación (haga clic en la pestaña "Exportar puntaje para EdX") y péguelo en el calificador de tareas en edX para obtener su calificación del curso.

## Formato de envío

El formato para el archivo de envío es dos columnas con `row_id` y `accepted`. El tipo de datos de `accepted` es un entero, solo los valores válidos son 0 y 1.

Si predice que se 1 acepta para cada encuestado, el `.csv` archivo que envíe tendrá el siguiente aspecto:

```
row_id,accepted
0,1
1,1
2,1
3,1
4,1
⋮
```

## Métrica de rendimiento

GOVERNMENT-  
DATA/LEADERBOARD/)

DESCARGA DE DATOS  
(/COMPETITIONS/14/MORTGAGE-  
APPROVALS-FROM-  
GOVERNMENT-DATA/DATA/)

ENVÍOS 4  
(/COMPETITIONS/14/MORTGAGE-  
APPROVALS-FROM-  
GOVERNMENT-  
DATA/SUBMISSIONS/)

REGLAS OFICIALES  
(/COMPETITIONS/14/MORTGAGE-  
APPROVALS-FROM-  
GOVERNMENT-DATA/RULES/)

Estamos prediciendo una variable binaria, por lo que este es un problema de clasificación. Para medir la clasificación, utilizaremos una métrica conocida como precisión (también conocida como "tasa de clasificación"). La precisión es una medida para evaluar los modelos de clasificación. Informalmente, la precisión es la fracción de predicciones que nuestro modelo acertó. Para la clasificación binaria, la precisión se puede calcular en términos de positivos y negativos de la siguiente manera:

$$\text{accuracy} = (TPAG + Tnorte) / (TPAG + Tnorte + FPAG + Fnorte)$$

donde TP significa verdaderos positivos, TN significa verdaderos negativos, FP significa falsos positivos y FN significa falsos negativos.

## Características

Hay 21 variables en este conjunto de datos. Cada fila del conjunto de datos representa una solicitud de préstamo informada por la HMDA, y el conjunto de datos con el que estamos trabajando cubre un año en particular.

Proporcionamos un identificador único llamado `lender` para cada institución de préstamo individual.

Las variables son las siguientes:

### UBICACIÓN DE LA PROPIEDAD

---

- `msa_md` (categórico): categórico sin orden que indica Área estadística metropolitana / División metropolitana donde un valor de `-1` indica un valor faltante
- `state_code` (categórico): categórico sin orden que indica el estado de EE. UU. donde un valor de `-1` indica un valor faltante
- `county_code` (categórico): categórico sin orden que indica el condado donde un valor de `-1` indica un valor faltante

## INFORMACIÓN DEL PRÉSTAMO

---

- `lender` (categórico): categórico sin pedido que indica cuál de los prestamistas fue la autoridad para aprobar o rechazar este préstamo
- `loan_amount` (int) - Tamaño del préstamo solicitado en miles de dólares
- `loan_type` (categórico): indica si el préstamo otorgado, solicitado o comprado fue convencional, garantizado por el gobierno o asegurado por el gobierno; Los valores disponibles son:

```
1 -- Conventional (any loan other than FHA, VA, FSA, or RHS loans)
2 -- FHA-insured (Federal Housing Administration)
3 -- VA-guaranteed (Veterans Administration)
4 -- FSA/RHS (Farm Service Agency or Rural Housing Service)
```

- `property_type` (categórico): indica si el préstamo o la solicitud fue para una vivienda de una a cuatro familias (que no sea vivienda manufacturada), vivienda manufacturada o vivienda multifamiliar; Los valores disponibles son:

```
1 -- One to four-family (other than manufactured housing)
2 -- Manufactured housing
3 -- Multifamily
```

- `loan_purpose` (categórico): indica si el propósito del préstamo o la solicitud fue para la compra de una vivienda, mejoras para el hogar o refinanciación; Los valores disponibles son:

```
1 -- Home purchase
2 -- Home improvement
3 -- Refinancing
```

- `occupancy` (categórico): indica si la propiedad a la que se refiere la solicitud de préstamo será la vivienda principal del propietario; Los valores disponibles son:

```
1 -- Owner-occupied as a principal dwelling
2 -- Not owner-occupied
3 -- Not applicable
```

- `preapproval` (categórico): indique si la solicitud o el préstamo incluyeron una solicitud de aprobación previa de un préstamo para la compra de una vivienda; Los valores disponibles son:

```
1 -- Preapproval was requested
2 -- Preapproval was not requested
3 -- Not applicable
```

## INFORMACIÓN DEL APLICANTE

---

- `applicant_income` (int) - En miles de dólares
- `applicant_ethnicity` (categórico) - Etnia del solicitante; Los valores disponibles son:

```
1 -- Hispanic or Latino
2 -- Not Hispanic or Latino
3 -- Information not provided by applicant in mail, Internet, or telephone ppl
  ication
4 -- Not applicable
5 -- No co-applicant
```

- `applicant_race` (categórico) - Raza del solicitante; Los valores disponibles son:

```
1 -- American Indian or Alaska Native
2 -- Asian
3 -- Black or African American
4 -- Native Hawaiian or Other Pacific Islander
5 -- White
6 -- Information not provided by applicant in mail, Internet, or telephone application
7 -- Not applicable
8 -- No co-applicant
```

- applicant\_sex (categórico) - Sexo del solicitante; Los valores disponibles son:

```
1 -- Male
2 -- Female
3 -- Information not provided by applicant in mail, Internet, or telephone application
4 or 5 -- Not applicable
```

- co\_applicant (bool): indica si hay un cosolicitante (a menudo un cónyuge) o no

## INFORMACIÓN DEL CENSO

---

- population - Población total en el tracto
- minority\_population\_pct - Porcentaje de la población minoritaria con respecto a la población total para el tracto
- ffiecmedian\_family\_income - Ingresos familiares medios de FFIEC en dólares para el MSA / MD en el que se ubica el tracto (ajustado anualmente por FFIEC)
- tract\_to\_msa\_md\_income\_pct -% del ingreso familiar medio del tracto en comparación con el ingreso familiar medio de MSA / MD
- number\_of\_owner-occupied\_units - Número de viviendas, incluidos condominios individuales, en las que vive el propietario.
- number\_of\_1\_to\_4\_family\_units - Viviendas construidas para albergar a menos de 5 familias.

## ÍNDICE Y VARIABLE OBJETIVO

---

- `row_id` - Un identificador único sin significado intrínseco, pero las ID en su envío deben coincidir exactamente con el formato de envío
- `accepted` - Indica si la solicitud de hipoteca fue aceptada (originada con éxito) con un valor de 1 o rechazada con un valor de 0

## Fila de ejemplo

Aquí hay un ejemplo de una de las filas en el conjunto de datos para que pueda ver los tipos de valores que puede esperar en el conjunto de datos. La mayoría son categóricos, algunos son numéricos y pueden faltar valores.

<code>row_id</code>	0 0
<code>tipo_préstamo</code>	1
<code>tipo de propiedad</code>	1
<code>propósito del prestamo</code>	3
<code>ocupación</code>	1
<code>monto del préstamo</code>	116
<code>aprobación previa</code>	3
<code>msa_md</code>	24
<code>código del estado</code>	4 4
<code>código_condado</code>	106
<code>etnia del solicitante</code>	2
<code>raza solicitante</code>	3
<code>solicitante_sex</code>	2
<code>Ingresos del solicitante</code>	66
<code>población</code>	3263
<code>minority_population_pct</code>	52,815
<code>ffiecmedian_family_income</code>	71852

row_id	0 0
trate_to_msa_md_income_pct	81,198
número_de_unidades_ocupadas por el propietario	786
number_of_1_to_4_family_units	1067
prestador	494
co_aplicante	Falso

## Referencias

- FFIEC. Ley de divulgación de hipotecas para el hogar <https://www.ffiec.gov/hmda/default.htm> ().
- Hoja de código de (<https://www.ffiec.gov/hmdarawdata/FORMATS/2015HMDACodeSheet.pdf>) registro de solicitud de préstamo de HMDA .

Términos de uso (/termsofuse/) · Política de derechos de autor (/copyrightpolicy/) · Política de privacidad (/privacypolicy/)

Copyright © 2019 DrivenData, Inc. Microsoft® y el logotipo de Microsoft son marcas comerciales de Microsoft, Inc. EdX® y el logotipo de edX son marcas comerciales o marcas comerciales registradas de edX Inc. Todos los derechos reservados.

Trabaja con nuestra comunidad (<https://www.drivendata.org/>) · Trabaja con nuestro equipo (<http://www.drivendata.co/>)

Alojado por (<https://www.drivendata.org/>)