

Introduction to the Florida 2013 Dataset

This dataset is derived from the Florida 2013 Home Mortgage Disclosure Act (HMDA) data and contains detailed information about mortgage loan applications. It provides insights into applicant demographics, loan characteristics, and outcomes. The dataset includes numeric and categorical features, covering various aspects of mortgage applications in Florida for the year 2013.

Below is an explanation of each column in the dataset:

Column Descriptions

1. **ID** (*int*)
2. **agency** (*object*)
The regulatory agency responsible for overseeing the financial institution that processed the loan application. Examples include *Consumer Financial Protection Bureau* or *National Credit Union Administration*.
3. **loan_type** (*object*)
The type of loan requested. Examples include *Conventional* or *FHA-insured* loans.
4. **property_type** (*object*)
Specifies the type of property being financed, such as *One-to-four family dwelling* or *Manufactured home*.
5. **loan_purpose** (*object*)
Indicates the purpose of the loan, such as *Home purchase*, *Home improvement*, or *Refinancing*.
6. **owner_occupancy** (*object*)
Denotes whether the property is occupied by the owner as a principal residence or rented out. Categories include *Owner-occupied* or *Not owner-occupied*.
7. **loan_amount** (*int64*)
The amount of the loan in thousands of dollars.
8. **preapproval** (*object*)
Indicates whether the loan application was pre-approved or if preapproval was not applicable.
9. **msamd** (*object*)
The Metropolitan Statistical Area/Metropolitan Division (MSA/MD) name where the property is located.
10. **county** (*object*)
The county where the property is located.
11. **county_code** (*float64*)
A numerical code representing the county.

12. **census_tract_number** (*float64*)
A geographical identifier representing a subdivision of the county used for demographic analysis.
13. **applicant_ethnicity** (*object*)
The ethnicity of the loan applicant, such as *Hispanic or Latino* or *Not Hispanic or Latino*.
14. **applicant_race_name_1, applicant_race_name_2, applicant_race_name_3** (*object*)
The racial identity of the applicant, with up to three reported race categories. Examples include *White*, *Asian*, or *Black or African American*.
15. **co_applicant_ethnicity** (*object*)
Ethnicity of the co-applicant (if any). Categories are similar to the applicant's ethnicity.
16. **co_applicant_race_name_1, co_applicant_race_name_2, co_applicant_race_name_3** (*object*)
Racial identity of the co-applicant, with up to three reported race categories.
17. **applicant_sex** (*object*)
The gender of the applicant, such as *Male* or *Female*.
18. **co_applicant_sex** (*object*)
The gender of the co-applicant.
19. **applicant_income** (*float64*)
The applicant's reported income in thousands of dollars.
20. **lien_status** (*object*)
Indicates the lien status of the loan, such as *Secured by a first lien* or *Not secured by a lien*.
21. **population** (*float64*)
The population of the census tract where the property is located.
22. **minority_population** (*float64*)
The percentage of the population in the census tract that belongs to a minority group.
23. **hud_median_family_income** (*float64*)
The median family income for the census tract, as defined by the Department of Housing and Urban Development (HUD).
24. **tract_to_msamd_income** (*float64*)
The ratio of the median family income of the census tract to the median family income of the MSA/MD.
25. **number_of_owner_occupied_units** (*float64*)
The number of owner-occupied housing units in the census tract.
26. **A, B, C, D – anonymous columns** – could be useful for prediction

27. **label** (*object*)

The outcome of the loan application process. Examples include *Loan originated* or *Application denied by financial institution*.