Formalized Project Proposal

Rob Osborne

2025-03-04

1. Research Question

Clearly state your final research question, ensuring it is well-defined and feasible within the scope of the project.

Question

What factors in pet adoption listings influence adoption speed and success?

Other topics and questions of interest:

- Which pet profile attributes (e.g., breed, age, size, color, health status) are most predictive of faster adoption?
- Do certain breeds or pet types (dogs vs. cats) have a significantly higher or lower adoption rate?
- How do adoption rates vary by location, shelter type, or listing details? "What role do pet descriptions and images play in influencing adoption likelihood?"
- Are there common characteristics among pets that remain listed for extended periods without adoption?
- What organizations or shelters have the highest adoption rates, and what factors contribute to their success?

2. Data Collection Methods (Ingestion Phase)

Identify the data sources you plan to use (e.g., API, web scraping, or openly available datasets) and provide links if available.

Describe how you will ingest the data into a storage solution, including any planned automation or scheduled tasks.

Data Sources

APIs

- Petfinder API
- Rescue Groups API

Datasets

- Sonoma county intake/outtake data (Also has an API)

Data Ingestion

I plan to create a DB through Supabase and create several edge functions that will run on a schedule to fetch data from the listed API.

These functions will do one of the following to the provided APIs: - Fetch most up to date pets that have been adopted. - Fetch pets that are still listed as adoptable after 60 days of lambda invocation.

For the datasets, I plan to run functions locally once to transform and inject into the DB.

Why am I choosing Supabase?

- They have a free tier.
- Out of the box, the DB you create will generate an API for you.
- They have support for scheduling edge functions which will be needed to fetch data.

Note: I wanted to create this project using AWS for scalability and practice creating and managing an RDS instance, but Supabase will handle the data serving part out of the box.

3. Data Transformation Process

Outline how you will structure and normalize the raw data into a database schema (preferably in third normal form (3NF)).

Explain how you will clean, preprocess, and transform the data for later analysis.

Database Schema

Animals Stores general information about the animals.

Column Name	Data Type	Constraints	Description
`id`	BIGINT	PRIMARY KEY	Unique ID for each animal
`name`	VARCHAR(255)	NOT NULL	Animal's name
`age`	ENUM	NOT NULL	kitten, adult, senior, unknown
`species`	ENUM	NOT NULL	dog, cat
`breed`	VARCHAR(100)	UNIQUE	Breed name (e.g., Rottweiler)
`sex`	VARCHAR(10)	NOT NULL	Male/Female
`size`	ENUM	NOT NULL	small, medium, large
`description`	TEXT	NULL	Animal description
`status`	ENUM	NOT NULL	available, adopted, pending
`organization_id`	VARCHAR(50)	FOREIGN KEY	References `organizations.id`
`posting_img_count`	SMALLINT	NULL	The count of the images in the posting

Attributes Stores various attributes of animals.

Column Name	Data Type	Constraints	Description
`animal_id`	BIGINT	FOREIGN KEY	References `animals.id`
`spayed_neutered`	BOOLEAN	NULL	Whether the pet is sterilized
`house_trained`	BOOLEAN	NULL	House trained or not
`declawed`	BOOLEAN	NULL	Only relevant for cats
`special_needs`	BOOLEAN	NULL	Special needs flag
`shots_current`	BOOLEAN	NULL	Whether vaccinations are current

Environment Stores compatibility of an animal with other pets and children.

Column Name	Data Type	Constraints	Description
`animal_id`	BIGINT	FOREIGN KEY	References `animals.id`
`dogs_ok`	BOOLEAN	NULL	Whether the pet is good with dogs
`cats_ok`	BOOLEAN	NULL	Whether the pet is good with cats
`kids_ok`	BOOLEAN	NULL	Whether the pet is good with children

Organizations Stores information about the organizations managing adoptions.

Column Name	Data Type	Constraints	Description
`id`	VARCHAR(50)	PRIMARY KEY	Unique organization ID
`name`	VARCHAR(255)	NOT NULL	Organization name
`city`	VARCHAR(255)	NOT NULL	Operating city
`state`	VARCHAR(50)	NOT NULL	Operating state

Data Transformation

Notes about data transformation: - Missing values will be inserted with `NULL`. - Responses from the API will be transformed into the appropriate data types. - Data will be processed through separate scheduled functions. - Dataset will be processed one time on my local machine.

4. Data Serving and Querying

Mention any plans to expose data through an API, dashboard, or other means if applicable.

I plan to expose the data through an API generated from Supabase.