BA820 – Project Proposal

Cover Page

- **Project Title: Understanding the Living Habits of Pet Cats in Cornwall, UK**
- **Section and Team Number: B1-Team 12**
- **Members: Fei Han, Kaixin Gao, Kang Ni, Xiaoyan Wang**

1. Primary Dataset

1.1. Project Motivation and Preliminary Exploratory Data Analysis (EDA)

Domain and business context

Pet cats are widely distributed across urban areas, yet their living habit is often treated as homogeneous despite growing evidence of substantial individual variation. Analysis of the Pet Cats UK dataset shows that while most cats remain close to their homes, a small subset exhibits irregular living habits. Understanding these differences is important because cats' movement behaviors directly influence ecological impacts, exposure to risk, and interactions with humans and the urban environment. By identifying distinct movement patterns among urban pet cats, this project aims to support more targeted and evidence-based decision-making for conservation groups, policymakers, and pet owners.

Stakeholders & Decision-Makers

Veterinarians, Pet owner, Ecologist, Urban biodiversity–focused conservation organizations Urban planners, and local councils

1.2. Domain / Business Questions

Q1: How can unsupervised machine learning methods be used to identify spatial clustering patterns in cat locations, and how do these clusters relate to environmental factors and observed behavioral traits?
Insights: Based on the distribution map in the EDA section, clear spatial clusters are observed, motivating further analysis of the factors driving these clusters and the impacts they may have.
Significance: This analysis is important for understanding cats' living conditions and spatial distribution across urban areas. Policymakers and city councils can use these insights to develop more effective and region-specific animal control policies. Biodiversity conservation groups can also leverage this information to guide member allocation and resource deployment.

Q2: How do cats form distinct clusters when analyzed by time period and ground speed using unsupervised methods?

Insights: The EDA results show that cats exhibit particularly high activity levels during nighttime hours, especially around 11 PM and 12 AM. Analyzing temporal activity patterns provides deeper insight into cats' living habits and behavioral dynamics.

Significance: Understanding cats' daily activity patterns helps researchers examine behavioral traits and environmental influences, while enabling city governments to develop targeted, time-based strategies to reduce animal-related damage and improve urban coexistence.

Q3: How do male cats and female cats distribute across the city, and how do their living traits influence differences in these spatial patterns?

Insights: The distribution of animal reproductive condition in EDA part shows that female cats are substantially underrepresented compared to male cats. This pronounced gender imbalance in

the dataset was unexpected and motivated further investigation into the potential impact of cat gender on observed behavioral patterns.

Significance: This analysis is important for potential pet owners considering the adoption of a cat. It helps evaluate the advantages and disadvantages of female and male cats by comparing individual behavioral traits and assessing their potential impacts on the ecological system.

Q4: When cats are clustered based on overall living habits—such as activity regularity and spatial concentration around the home—do distinct behavioral profiles emerge, and how do different feeding types distribute across these groups?

Insights: The Food Type Consumption graph illustrate the food habits for each cat, most of the cats will have both dry and wet food, some of them have Dry only, and very few will have wet only. We do think this is a very interesting part, and we can work on it.

Significance: This analyzes the Information on feeding food types can help government animal control agencies and pet owners better understand how feeding behaviors influence cats' living patterns. These insights can also support animal shelters and pet owners in selecting appropriate food types to protect biodiversity and ecological systems while improving interactions between cats and humans.

2. Backup Dataset

2.1 Preliminary Exploratory Data Analysis (EDA)

Domain and business context

The Hollywood Age Gap dataset captures age differences between on-screen couples in films released between 1935 and 2022. Each observation represents a paired relationship between two actors as partners. It includes their respective ages at the time of release, gender, and the age gap. This dataset covers decades of changes in movie characters and the evolving expectations of the film industry and audiences regarding romantic relationships.

Stakeholders & Decision-Makers

Film and television production companies, streaming platforms, actors and their management teams, filmmakers and casting teams.

2.2. Domain / Business Questions

Q1: Can film couples be grouped into common age patterns, and what differences exist in their age gaps and pairing directions?

Insights: The age combinations of couples in the film are not random. The age difference distribution shows that certain age pairings are significantly more common. Even after splitting by gender, different age differences and pairing directions can be observed, indicating the presence of recurring age pairing patterns in the film.

Significance: These findings can help film production and casting teams understand common age pairing patterns. For content platforms, they can also help differentiate between mainstream age combinations and relatively niche or higher-risk combinations.

Q2: Do age-gap patterns differ when the older partner is male versus female, and can these be treated as two distinct relationship types?

Insights: The age difference and pairing patterns differ depending on whether the older partner is male or female. Therefore, these two types of relationships can be considered two distinct relationship types.

Significance: For film and television companies, this helps in determining whether existing works are overly reliant on a particular gender narrative pattern. For platforms and production companies, this analysis can be used for content diversity assessment and to provide data support for projects related to female perspectives.

Q3: Does the narrowing age gap in film couples reflect a real shift in industry attitudes, or is it mainly driven by changes in film genres?

Insights: The average age gap between film couples has become smaller in recent years. However, this trend may be driven by changes in film genre proportions rather than a full shift in industry attitudes. As a result, traditional age-pairing patterns still remain in some genres.

Significance: This helps decision-makers judge whether the smaller age gap comes from changes in film genres or a real shift in industry attitudes. This is important for long-term strategy, as it helps decide whether to follow the mainstream or create different content.

Q4:Are extreme age-gap relationships in films isolated cases, or recurring patterns linked to specific actors or character types?

Insights: The extreme age gaps in film relationships are not accidental, but rather a recurring pattern. These relationships are often associated with specific actors or fixed character types, indicating that large age differences are consistently used.

Significance: This helps filmmakers determine if actors are repeatedly cast in roles with a significant age gap, and whether this setup has become a commercial formula. For agents and casting managers, it also helps actors and their teams assess whether there is a problem of typecasting in their roles.
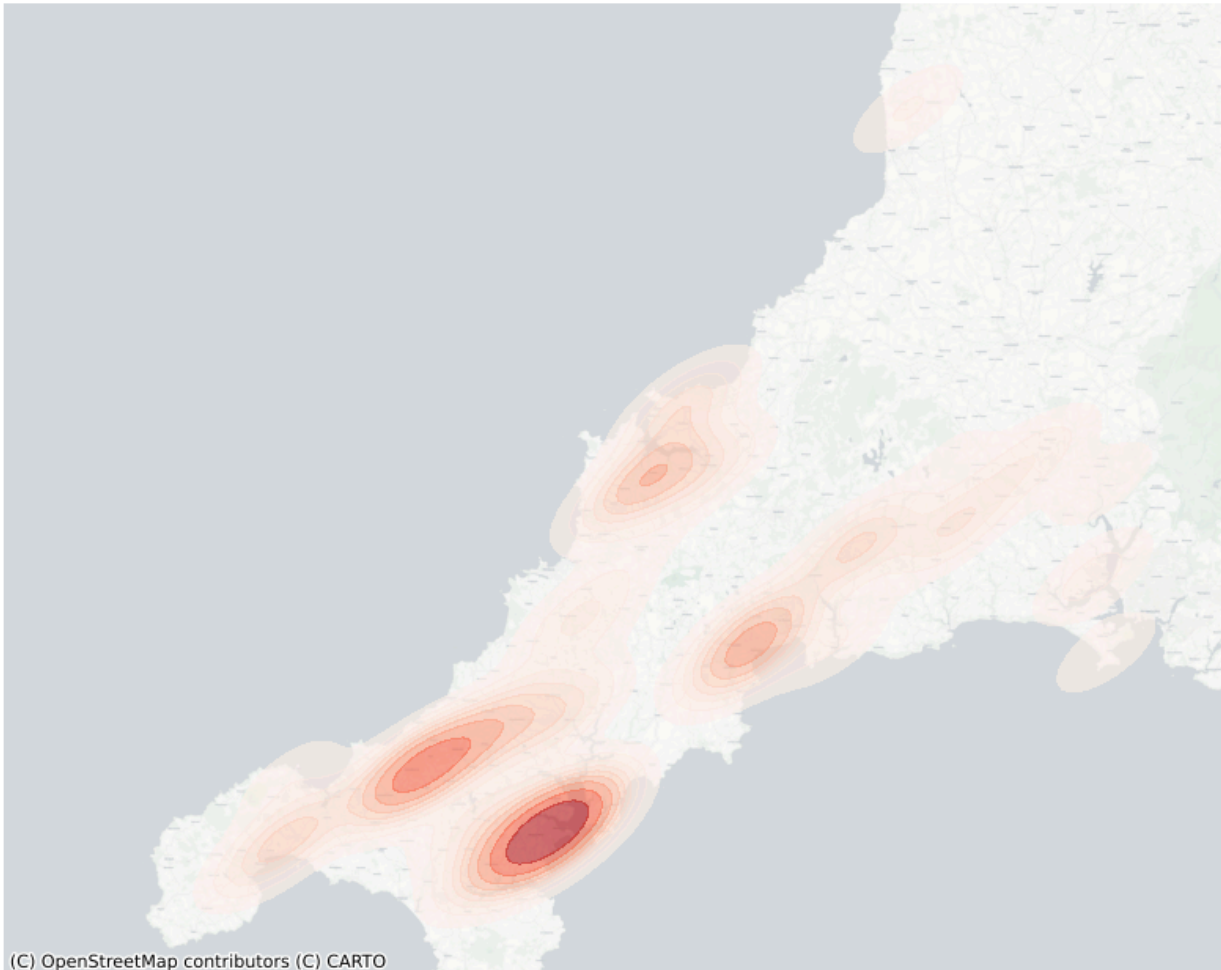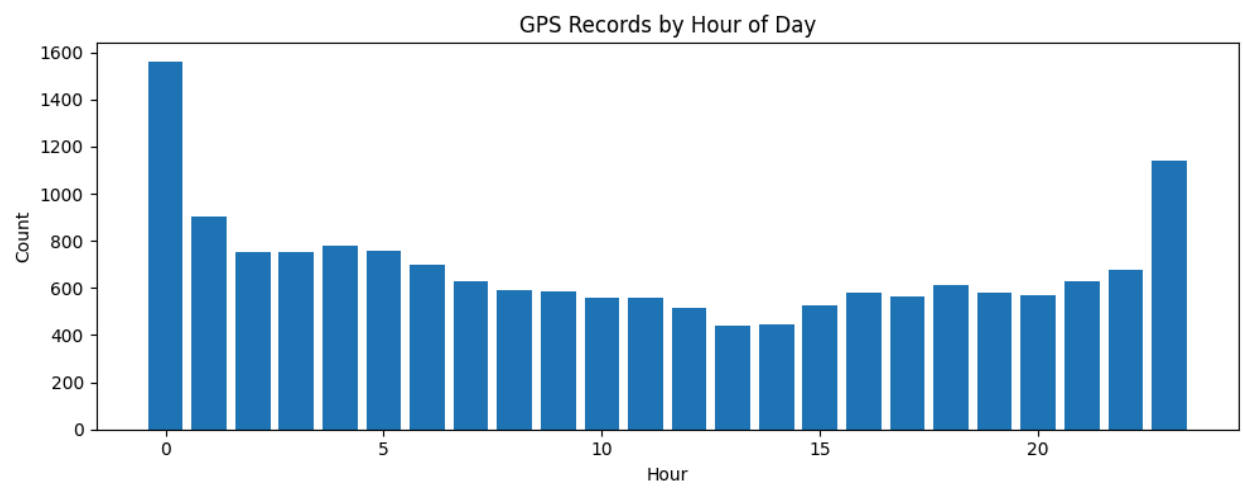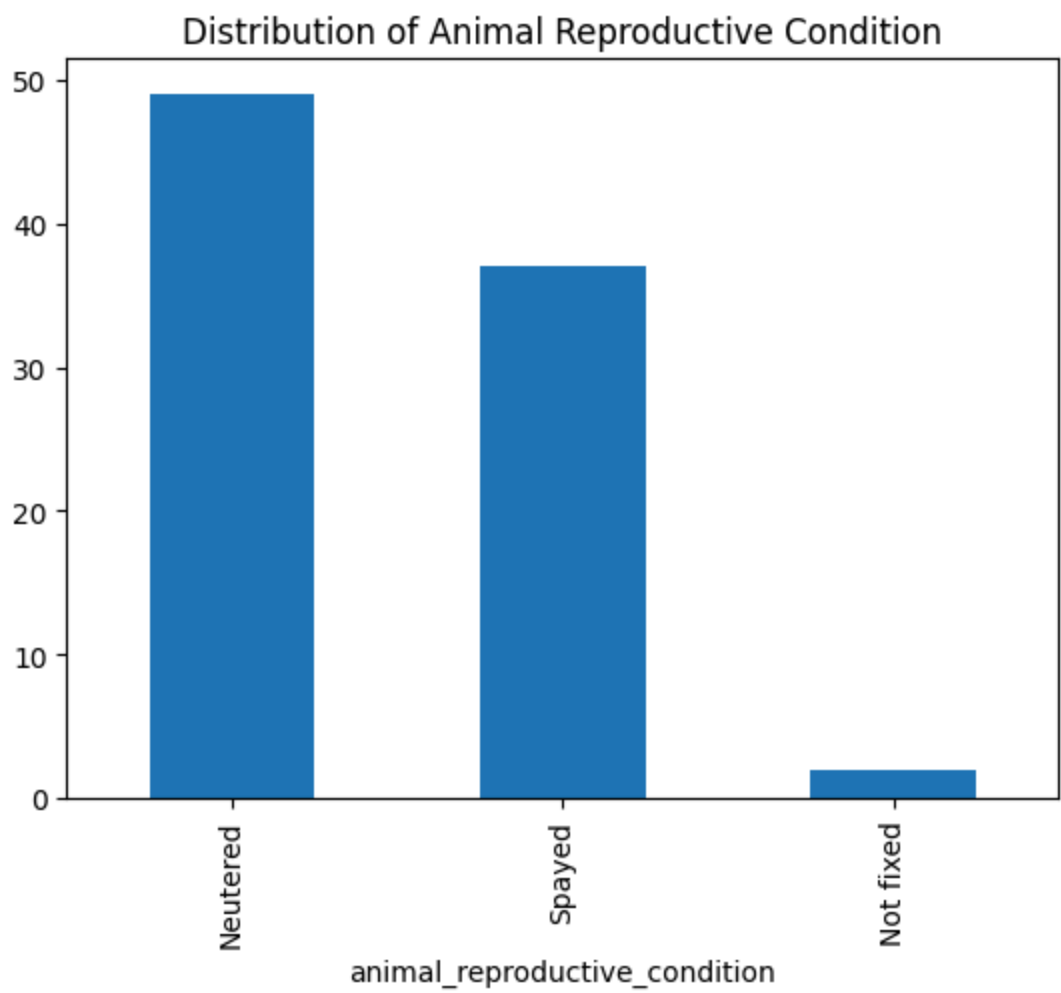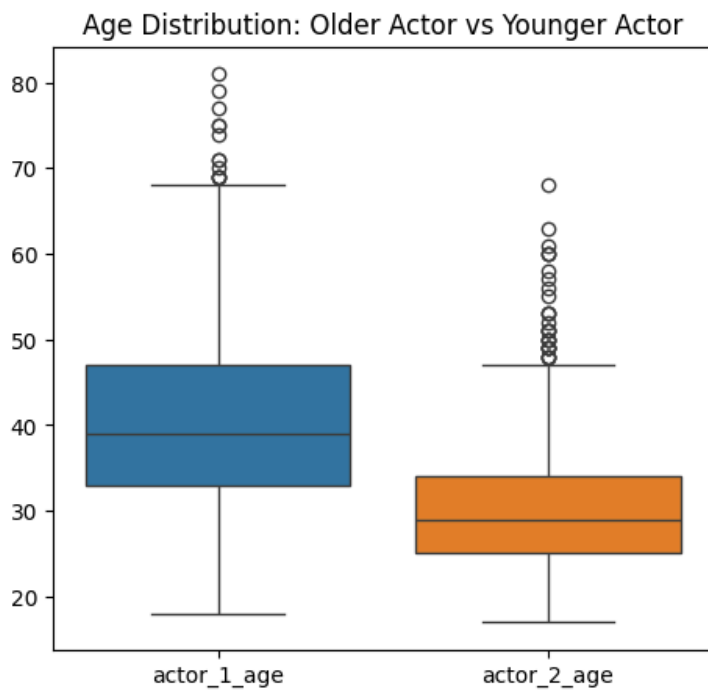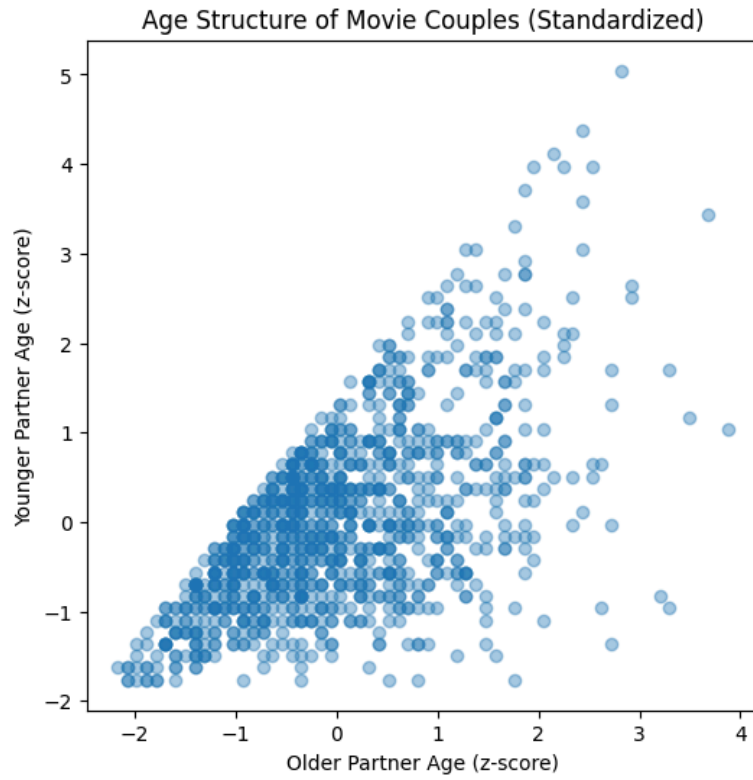
4. Appendix

GitHub Repository (Required)
Provide a link to your team's shared GitHub repository. This repository will serve as the central workspace for the semester-long project. The repository should include any EDA code. Your repository should allow your team to collaborate consistently, and its commits must maintain a clean record of work over time.
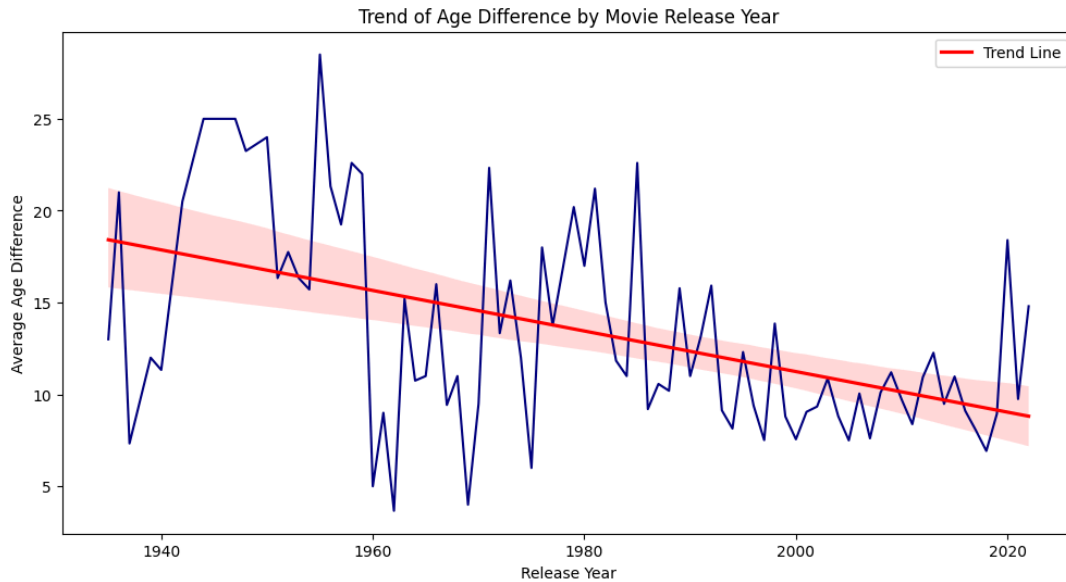
Supplementary Material (Highly Recommended)

Spatial Heatmap with Real Map Background

Distribution of Animal Reproductive Condition

GPS Records by Hour of Day

Age Structure of Movie Couples (Standardized)

Age Distribution: Older Actor vs Younger Actor

Trend of Age Difference by Movie Release Year

Process Overview

- Data collection and integration of movement and reference data
- Data cleaning and basic preprocessing
- Exploratory analysis to identify variability and patterns
- Discovery of natural behavior profiles and extreme cases
- Interpretation of findings in a domain and decision-making context

External References and Context

Kays, Roland, et al. "The Small Home Ranges and Large Local Ecological Impacts of Pet Cats." *Animal Conservation*, vol. 23, no. 5, 2020, pp. 516–526, https://doi.org/10.1111/acv.12563.

McDonald, Justin L., and Cole H. "Data from: The Small Home Ranges and Large Local Ecological Impacts of Pet Cats [United Kingdom]." *Movebank Data Repository*, 2020, https://doi.org/10.5441/001/1.pf315732.

Use of Generative AI Tools

- Coding Aid(part)
- Text Revision
- Grammar Check
- Project Initial Brainstorm
- Formating

ChatGPT Link:

https://chatgpt.com/share/697a9b46-3ef8-8010-bee1-777f61fbd716

https://chatgpt.com/share/697a9d19-672c-8013-8590-1e9d1a087f95

https://chatgpt.com/share/697a9a8d-28f0-800e-83dc-1ab731693bee