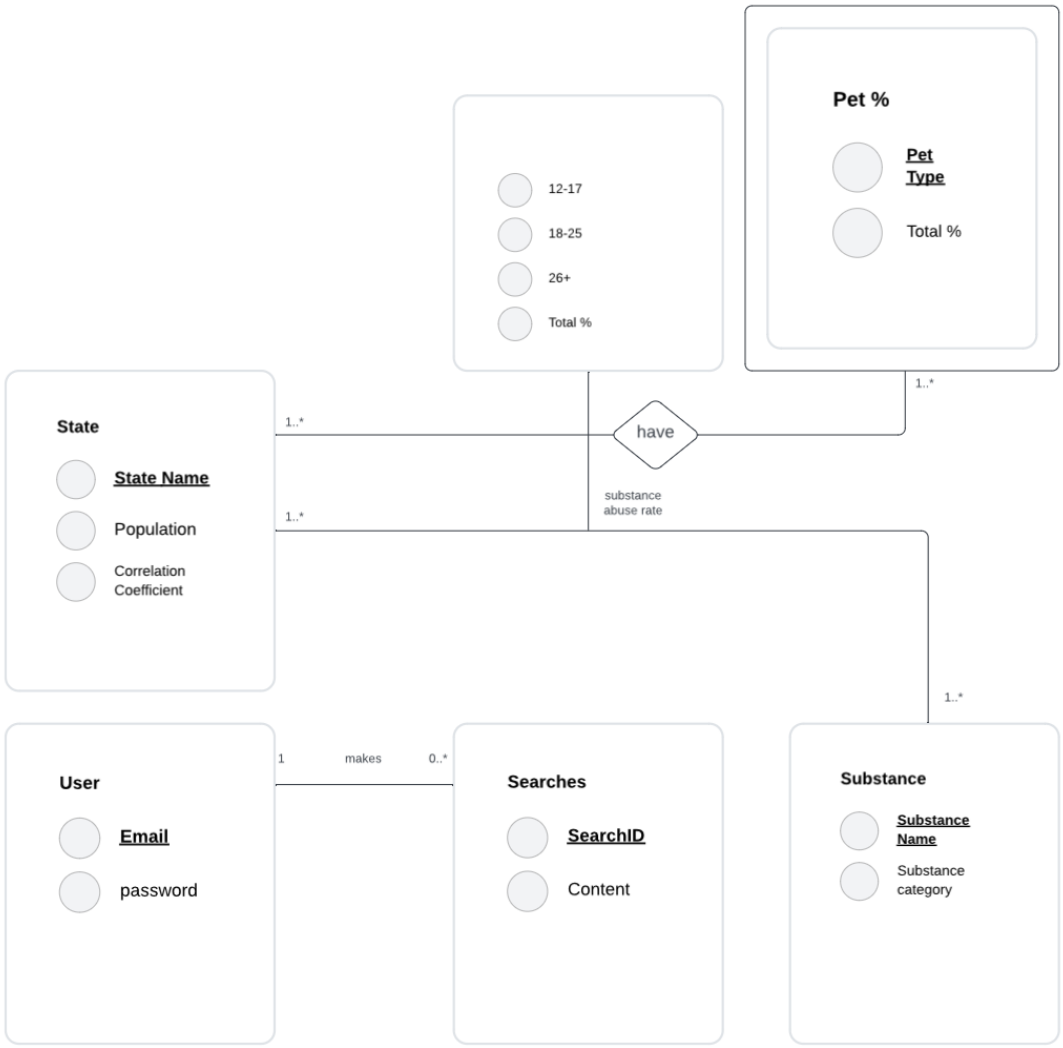
## UML Diagram




## Assumptions

- State - Pet, **many to many**
    - Multiple pets correspond to multiple states
    - (Dog, Cat) – (AL, AK, AZ, AR, …)
- State - Substance, **many to many**
    - Multiple substances correspond to multiple states

    - (Meth, Heroin, Cocaine, …) – (AL, AK, AZ, AR, …)
- User → Searches, **one to many**
    - One user corresponds to multiple searches

- Substance Abuse Rate is a table that contains information regarding the relation between substance and state, not considered one of the five entities.

- User and Searches entities are included to implement a create and delete functionality.
    - User can make searches based upon certain parameters, information is saved.
    - Previous searches can be deleted.

**Normalization**

We chose to use Boyce-Code Normal Form (BCNF) to normalize our database specifically because we are dealing with multiple substances and there is a lot of overlap. For example, we could just take out the Substances table and make SubstanceName and SubstanceCategory two attributes in the Substance Abuse Rate table, but then it would be disorganized and StateName would be the only unique identifier. However, one state will have multiple different rates because there are several drugs/rates in a state, so in general, it wouldn't make sense for State to be the only unique identifier, substance has to be one too. Overall, we chose BCNF as it would help with organizing our complex datasets and correlations and having a more modulated, organized dataset was best for organizing all the different rates.

**BCNF Normalized Relation**

StateName → Population, CorrelationCoefficient
PetType → Total%, StateName
SubstanceName → SubstanceCategory
SearchID → Email, Content
Email → Password
SubstanceAbuseRate → Substance.SubstanceName, State.StateName

The above relation is normalized according to BCNF because each functional dependency in the relation adheres to the definition of BCNF which states that for any non-trivial functional dependency, the left hand side is a superkey of the relation. Since the relation was already BCNF normalized, no further simplification was required.

**Relational Schema**

State(State Name: VARCHAR(255) [PK], Population: INT, Correlation Coefficient: Decimal)

Pet %(Pet Type: VARCHAR(255)[PK], Total %:Decimal, State Name: VARCHAR(255)[FK])

Substance Abuse Rate(Substance Name: VARCHAR(255)[PK], State Name: VARCHAR(255)[PK], 12-17: Decimal, 18-25: Decimal, 26+: Decimal, Total %: Decimal)

User(Email: VARCHAR(255)[PK], password: VARCHAR(255))

Searches(SearchID: INT[PK], Email: VARCHAR(255) [FK], Content : VARCHAR(255))

Substance(Substance Name: VARCHAR(255)[PK], Substance category: VARCHAR(255))