CS513: Theory & Practice of Data Cleaning

**Final Project Phase 1 Report**

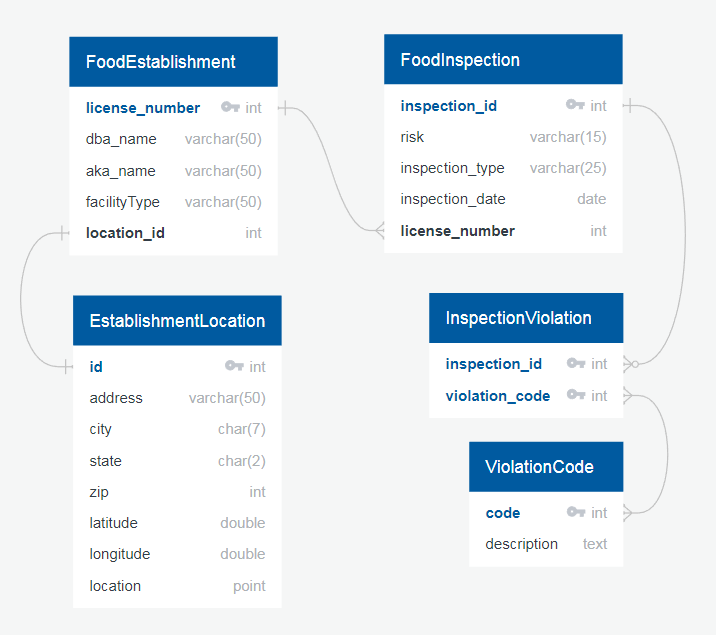
**Team 23**

Avinash Baldeo, <add Names>

[abaldeo2@illinois.edu](mailto:abaldeo2@illinois.edu), <add Emails>

1. **Dataset Overview**
2. **Dataset Description**

**2.1 Database Diagram**



**2.2 Database Schema**

**CREATE TABLE EstablishmentLocation (**  
 **id SERIAL PRIMARY KEY,**  
 **address VARCHAR(50),**  
 **city CHAR(7),**  
 **STATE CHAR(2),**  
 **zip INT,**  
 **latitude DOUBLE PRECISION,**  
 **longitude DOUBLE PRECISION,**  
 **location point**  
 **);**  
  
**CREATE TABLE FoodEstablishment (**  
 **license\_number INT PRIMARY KEY,**  
 **dba\_name VARCHAR(50),**  
 **aka\_name VARCHAR(50),**  
 **facilityType VARCHAR(50),**  
 **location\_id INT REFERENCES EstablishmentLocation(id)**  
 **);**  
  
**CREATE TABLE FoodInspection (**  
 **inspection\_id INT PRIMARY KEY,**  
 **risk VARCHAR(15),**  
 **inspection\_type VARCHAR(25),**  
 **inspection\_date DATE,**  
 **license\_number INT REFERENCES FoodEstablishment(license\_number)**  
 **);**  
  
**CREATE TABLE ViolationCode (**  
 **code INT PRIMARY KEY,**  
 **description TEXT**  
 **);**  
  
**CREATE TABLE InspectionViolation (**  
 **inspection\_id INT REFERENCES FoodInspection(inspection\_id),**  
 **violation\_code INT REFERENCES ViolationCode(code),**  
 **PRIMARY KEY (**  
 **inspection\_id,**  
 **violation\_code**  
 **)**  
 **);**

1. **Use cases**

**3.1 U0: Zero Cleaning Use Case**

**3.2 U1: Main Use Case**

**3.3 U2: Never Enough Use Case**

1. **Data Quality Problems**
2. **Phase-II Initial Plan**