**Healthcare Data Analysis - PEP Project 2**

*By Jasmine (Scrum Master), Ettione, Justin*

**Change History**

*Document created 2/5/24*

**Overview**

Respiratory illness is an ever-growing threat to the lives and health of the American people. Using existing data of past infections across the country, we will create a comprehensive analysis using data science and AWS technologies to discern patterns and potential predictions for future strategies in healthcare.

**Objectives**

* Create a working pipeline from file upload to cloud RDS
* Clean data via deployed Lambda functions
* Upload data to RDS
* Connect locally to RDS to generate analytics and extract insight from data patterns/trends
* Display analytics results on website hosted on EC2

**Success Metrics**

* Objectives met with minimal errors
* Successful data cleaning performed by AWS Lambda
* Successful import and export to/from cloud RDS
* Working EC2-hosted website

**Timeline & Release Planning**

Four days for Minimum Viable Product to be available for demonstration

**Personas**

This pipeline would be used to track and analyze respiratory illness infections to discern patterns and trends, which will be used as a basis for future healthcare strategies.

**Scenarios**

As a Data Scientist:

* + I want to explore the "Outpatient Respiratory Illness Activity Map" dataset to understand its structure, variables, and data quality.
  + I want to visualize temporal trends in respiratory illness activity over weeks and months to identify patterns.

As a Public Health Analyst:

* + I want to analyze regional variations in respiratory illness activity to identify hotspots and regions requiring targeted interventions.
  + I want to explore the correlation between severity levels and hospital admission rates to assess the impact on healthcare resources.

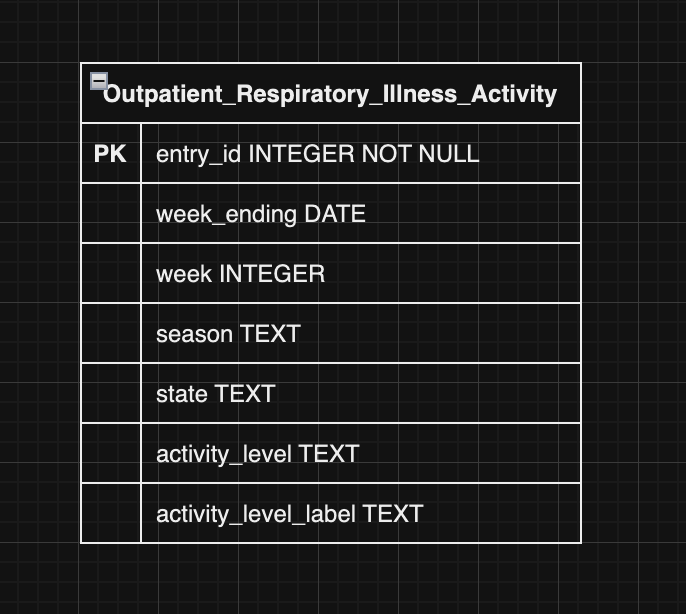
As a Data Engineer:

* I want to ensure seamless data transfer and storage by utilizing Amazon S3 to store and retrieve the "Outpatient Respiratory Illness Activity Map" dataset.
* I want to implement a data processing pipeline for continuous monitoring and updating of the analysis.

**Features**

* Parse and clean data from .csv file to export to cloud RDS
* Display results of data analytics on cloud-hosted website

**Database Diagram**

****