# **ETL PROJECT REPORT**

|  |  |
| --- | --- |
| **Topic:** | Home meal cooking or dining out |
| **Date:** | December 08, 2021 |
| **Document Authors:** | Kai Phan and Hoang-Yen Cao |
|  |  |

**Table of Contents**

[**ETL PROJECT REPORT** 1](#_Toc90199642)

[Objective 3](#_Toc90199643)

[Problem Statement 3](#_Toc90199644)

[Data Cleanup and Analysis 4](#_Toc90199645)

[Additional Links 5](#_Toc90199646)

# Objective

Provide a detailed analysis of the ETL(Extract-Transform-Load) process. We extract the data from Google Maps Platform from API, import it as a CVS file then load it into a database for storage.

# Problem Statement

Eating out is expensive and not healthy as homemade food. These analyses help readers decide to cook for their meals or dine at restaurants.

We collect healthy recipes from google and download restaurants' names from dataword.com.

# Data Cleanup and Analysis

We have performed the ETL process in three steps:

**Step1: Extraction**

We collected data from two sources:

1. The recipe data files used for extraction were from the public platform with the following link: https://airtable.com/appSDKSk3kO243A16/tblHJvDUEiZdgXfpX/viwzDiujtjtonzv26?blocks=hide
   1. We imported it into Pandas data frame in Jupyter notebook.
2. Restaurant data set: <https://maps.googleapis.com/maps/api/place/textsearch/json>

We loaded the restaurant’s location data was loaded into MongoDB(NoSQL).

Restaurant location dataset:

* 1. Using Python request to fetch the restaurant’s name, address.
  2. Executing printing out the data to keep track of the process.

**Step 2: Transformation**

We used two methods to transform the data.

1. Transform the CSV file. Recipe dataset: We used Python as the tool to perform the transformation on the recipe dataset using Panda’s library, described in the following steps:
   1. The Recipes-All Recipes. Cvs was first read into the Panda data frame, using read CVS.
   2. We dropped irrelevant data from the data frames by selecting the columns, Name, Rating, East of Prep, Type, Prep Time, Ingredients.
2. The restaurant data collection from Jason Request has been transforming during extraction and appending it into the list.

**Step 3: Loading**

This project type of data contains characters primarily. It does not require modifying. Therefore, we chose NoSQL for storage in this project.

We establish a connection Mongo Client for restaurant data, insert data in the database.

# Additional Links

1. Kai Pham GitHub Repository: <https://github.com/haipham0115/project-2__E.T.L>
2. HoangYen Cao Repository: https:// <https://github.com/ghhyc/ETL-Project>