**数据库设计**

**编制:雷邓君**

**项目名称:美国二手车估值系统**

**本文件修订历史信息:**

|  |  |  |  |
| --- | --- | --- | --- |
| **更新日期** | **修订者** | **批准人** | **主要修订摘要** |
| 2018.10.18 | 雷邓君 | 万可文 | 新编写文件 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

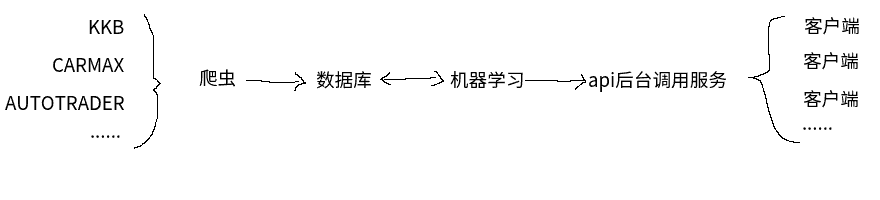
# 1.基础公共表设计

## 1.1项目目标

本项目是开发美国二手车价格评估系统.根据品牌，车型，款型，英里数，所在州等特征预测二手车的trade-in price(收购价),private party price(个人交易价),dealer price(车商零售价)及其中间价和范围。系统本身需具有可扩展性。

## 1.2项目总体架构

下图为估值系统level1结构图.目前以最简形式设计,后续根据项目进展再进行扩展.



# 2.项目子系统需求

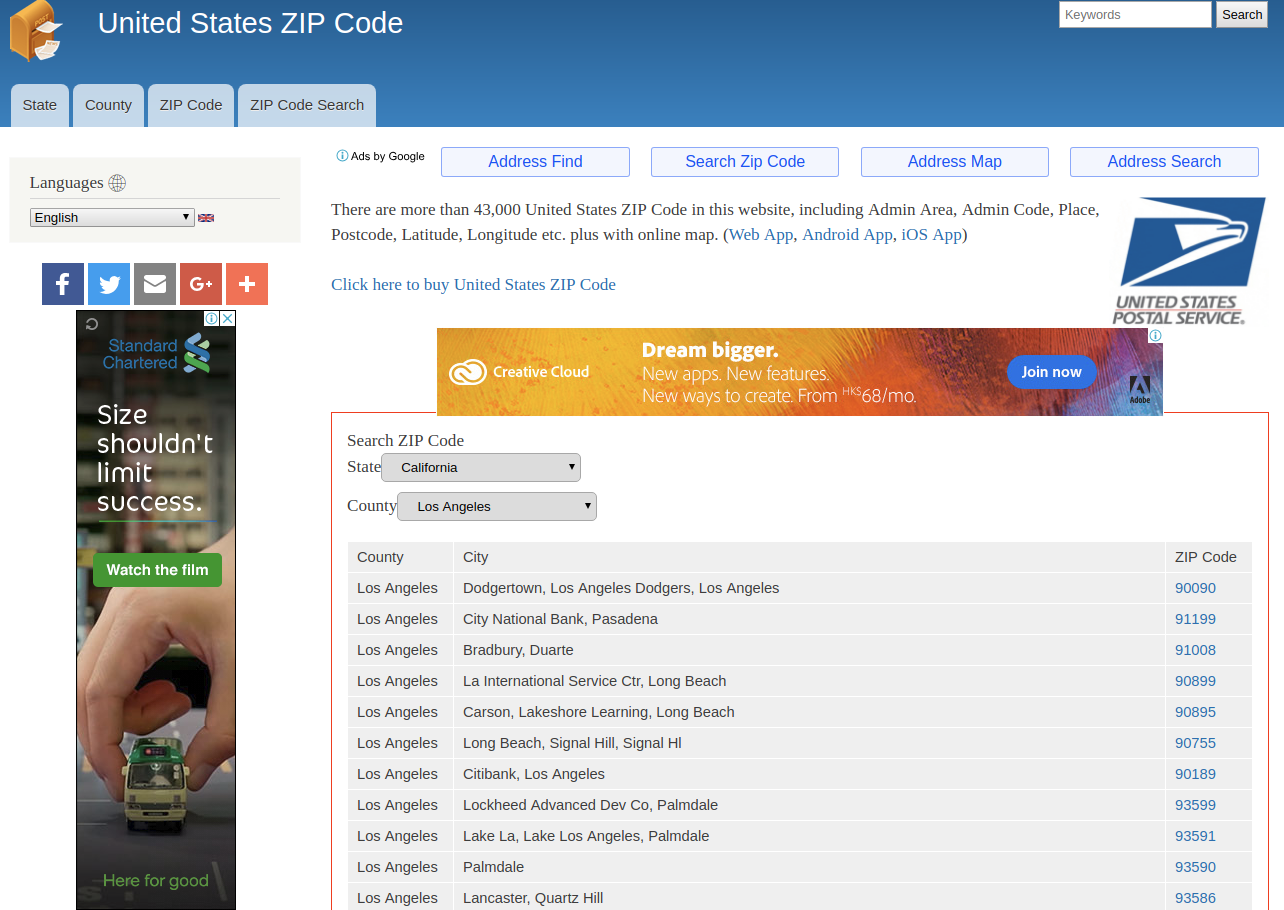
估值大系统分四个子系统,分别为爬虫系统,数据库系统,算法系统,api后端调用服务系统.

## 2.1爬虫系统需求

爬虫系统抓取二手车估值相关的平台的数据,并整理成标准数据录入到数据库,包括州/城市zip code,标准车型数据,平台车型数据,车源数据,竞争平台数据.

### 2.1.1 州/城市zip code表抓取需求

1.抓取state-county-city的zip code,建议抓取地址https://usa.youbianku.com/

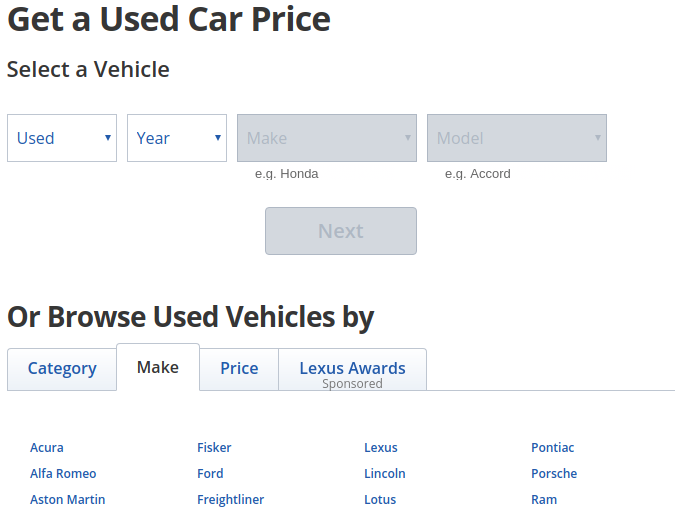


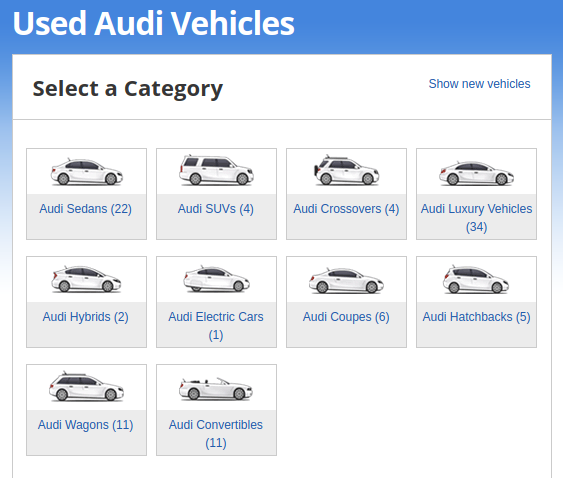
2.匹配zip code与lat long,建议地址下载https://simplemaps.com/data/us-zips

### 2.1.2标准车型库抓取需求

建议直接使用KBB的车型库作为我方的标准库,抓取包含以下内容:

1.所有的品牌,车系,款型,注意选择used(可看到所有历史车型),以下截图为抓取流程,优先级最高.

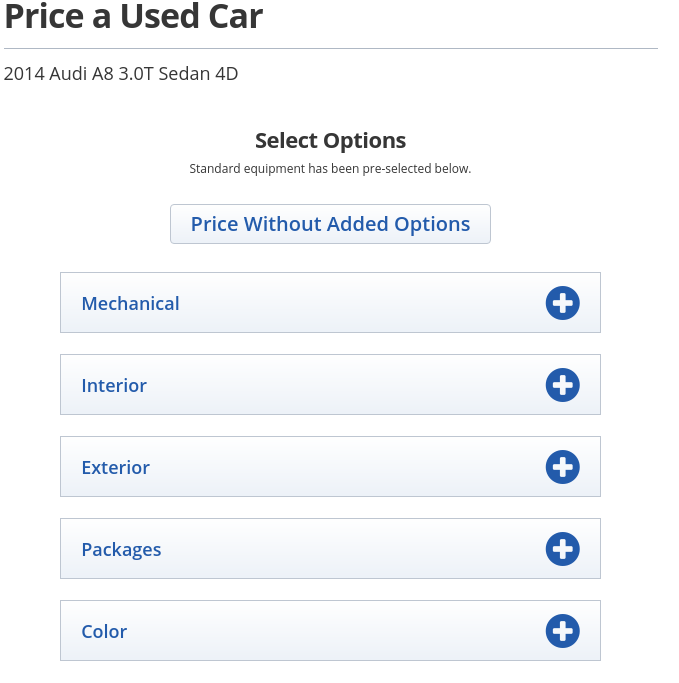


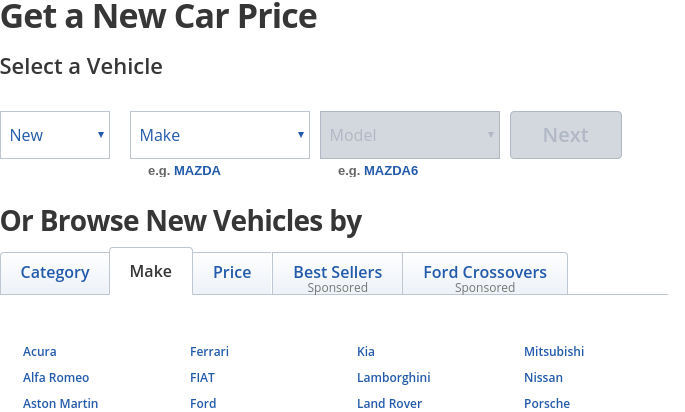




### 

### 

 2.参考1的抓取流程,抓取所有新车的特征,优先级最后.



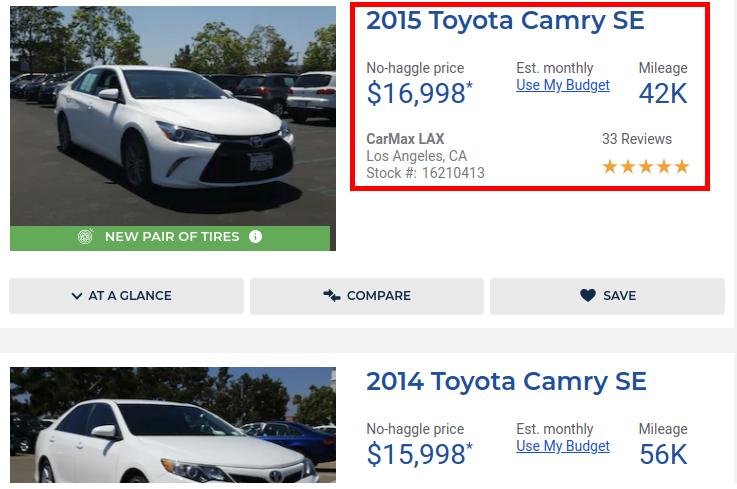
### 2.1.3 平台车型数据抓取需求

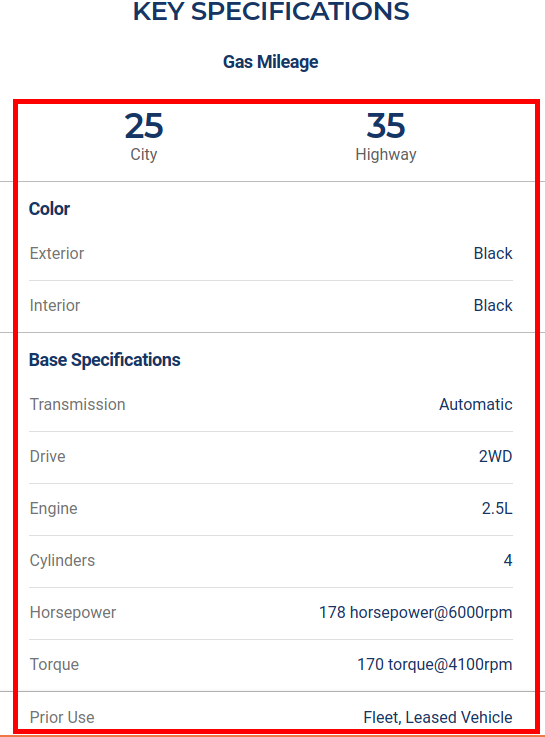
1.抓取carmax车型库

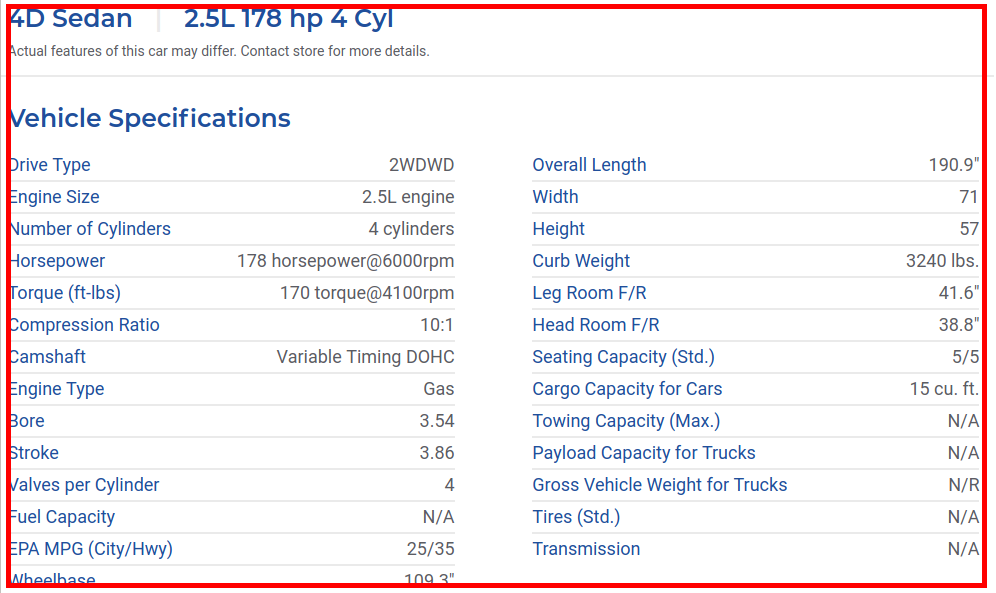
2.抓取autotrader车型库

### 2.1.4 平台车源数据抓取需求

1.抓取carmax车源数据,红框内为抓取内容



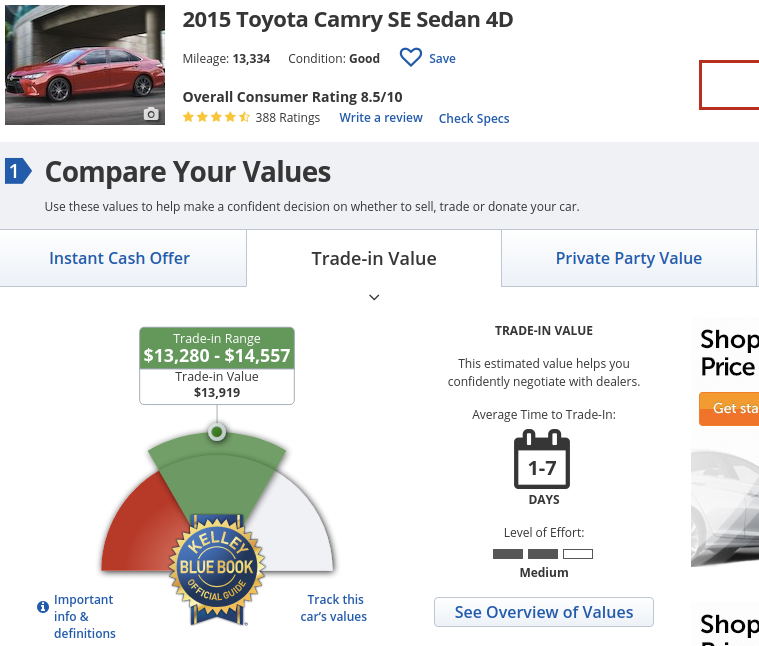


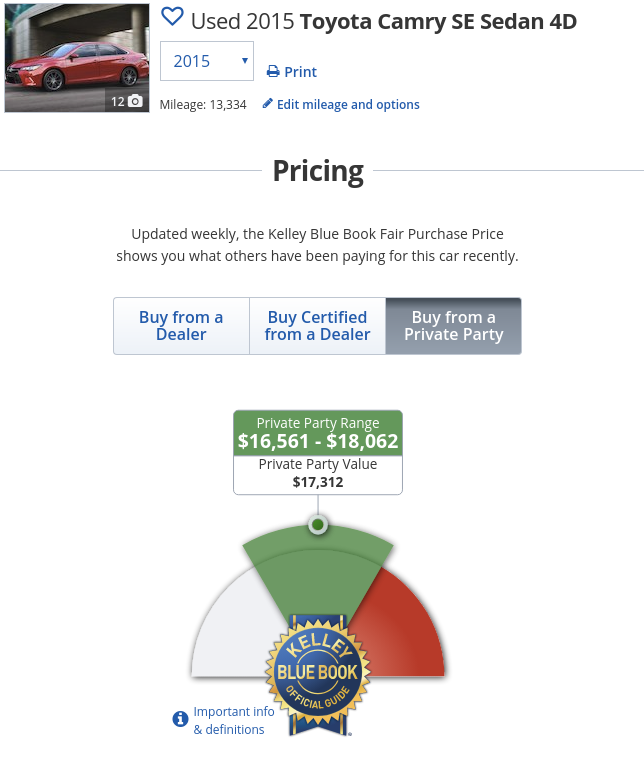


2.抓取autotrader车源数据,参考carmax抓取.

### 2.1.5 竞争平台估值抓取需求

1.根据算法组提供的车源数据,抓取kbb的四类价格估值,需要包含4类车况(fair,good,very good,excellent)





## 2.2 数据库系统需求

满足爬虫和算法系统对数据库的建表需求,并满足一定的性能需求.

## 2.3 算法系统需求

1.将各个平台的车型库匹配到我方标准车型库

2.将原始车源数据,通过算法匹配上品牌-车型-款型,并生成一张标准车源表

3.依据爬虫抓取数据,探索分析相关特征,训练生产模型,根据品牌-车型-款型-zip code-英里数-上牌年份预测二手车三类估值价格(收购价,个人交易价,车商零售价及其价格中间值和范围)

## 2.4 api后端调用服务需求

搭建http后端调用服务,通过发送附带参数(品牌-车型-款型-zip code-英里数-上牌年份)的http请求返回json格式的二手车估值三类价格.

目前若仅仅只做功能展示,可沿用公平价服务器,若要持续提供稳定服务给美国方合作伙伴,建议还是用亚马逊美国服务器.