第一题：海伦女士一直使用在线约会网站寻找适合自己的约会对象。尽管约会网站会推荐不同的任选，但她并不是喜欢每一个人。经过一番总结，她发现自己交往过的人可以进行如下分类：

* 不喜欢的人
* 魅力一般的人
* 极具魅力的人

数据中主要包含以下三种特征：

* 每年获得的飞行常客里程数
* 玩视频游戏所消耗时间百分比
* 每周消费的冰淇淋公升数

使用分类算法预测一下[56054,3.574967,0.494666]是不是喜欢的类型。

第二题：使用k-means算法，结合下面十个人的分数，分为学渣和学霸两类。

list1=[88,64,96,85]

list2=[92,99,95,94]

list3=[91,87,99,95]

list4=[78,99,97,81]

list5=[88,78,98,84]

list6=[100,95,100,92]

list7=[66,88,178,91]

list8=[80,97,60,92]

list9=[77,97,80,92]

list10=[99,95,88,100]

第三题：使用线性回归模型，进行房价预测

数据说明：

**时间和房价指数：housing\_price\_index.csv**

**失业指数：unemployment.csv**

字段：

total\_umemployed 完全没有工作的

more-than-15-weeks 超过15周的

not\_in... 没有找工作的

multi\_jobs 多份工作的

leavers 自动离职的

losers 被解雇的

**联邦政府贷款买房的利率：federal\_funds\_rate.csv**

**shiller.csv**

字段：

date 时间

sp500 500强公司股票指数

consumer\_price\_index 消费指数

long\_interest\_rate长期的利率

**工业生产总值：gross\_domestic\_product.csv**

字段：

total\_expenditures 消费

labor\_force\_pr 人力资源

producer\_price\_index 生产者价格指数

gross\_domestic\_product 国民生产总值

**第三题解题思路：**

1) 读取数据，并将数据按日期合并

2) 查看相关性

3) 建立模型