# 北京二手房数据分析与可视化

## 一、 学习目标[¶](file:///F:\\备课资料\\教学资料\\python\\python教学资料\\第二学期教学资料\\第6章%20数据分析与处理\\数据分析3天\\数据分析day01\\数据分析day01\\2%20code\\二手房数据分析案例.html" \l "1、-学习目标)

* 掌握数据处理的常用方法
* 知道数据分析的指标

**二、实现步骤**

**#1、导入所需包**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt #pandas自带的图形包

import os

from pyecharts.charts import Pie, Map, Bar, Line, Grid, Page #插件可以绘制很酷炫的图形

from pyecharts import options as opts

# matplotlib可以正常显示汉字

import matplotlib.pyplot as plt

plt.rcParams['font.sans-serif'] = ['Arial Unicode MS'] # 用来正常显示中文标签

plt.rcParams['axes.unicode\_minus'] = False # 用来正常显示负号

**#2、导入数据**

**#3、查看每列信息**

**#4、数据预处理**

**#4.1 初步处理**

**#（1）删除重复数据**

**#（2）去重**

# (3) 删除数据分析不需要的列。

•title：无需分析，删除

•detailUrl：无需分析，删除

•position：维度过细、删除

•houseInfo：提取室、厅、面积、方位、装修、楼层（高中低）、建筑年份、板塔

•followInfo：无需分析，删除

•tagInfo：提取是否靠近地铁

•totalPrice：提取房屋总价

•unitPrice：房屋单价

•county：无需处理

**# 4.2 提取信息（**houseInfo：提取室、厅、面积、方位、装修、楼层（高中低）、建筑年份、板塔**）**

#（1）提取室

#（2）提取厅

#（3）提取面积

#（4）提取方位朝向

#（5）提取装修类型

#（6）提取楼层

#（7）提取建筑年份

#（8）提取板塔

#（9）删除houseInfo

#（10）提取地铁

#（11）删除tagInfo列

#（12）查看删除后数据框信息

#（13）提取总价

#（14）提取单价

#（15）查看数据框信息

#（16）查看是否有空值

#（17）空值直接删除

#（18）转换数据类型（总价和单价转换为float）其他转换为整型

# (19) 查看前5条数据信息

#（20）查看数据框各列信息

#（21）查看方位朝向列

#（22）查看楼层列信息

#（23）将楼层列处理为：高层'、低层、中层、地下室'

#（24）将几个数据一般化（楼层、朝向、年份、板塔）

#（25）查看数据框各列信息

**#5、数据探索和可视化**

**#**5.1 北京二手房都处在哪些价位

#（1）设置分割点

#（2）按总价

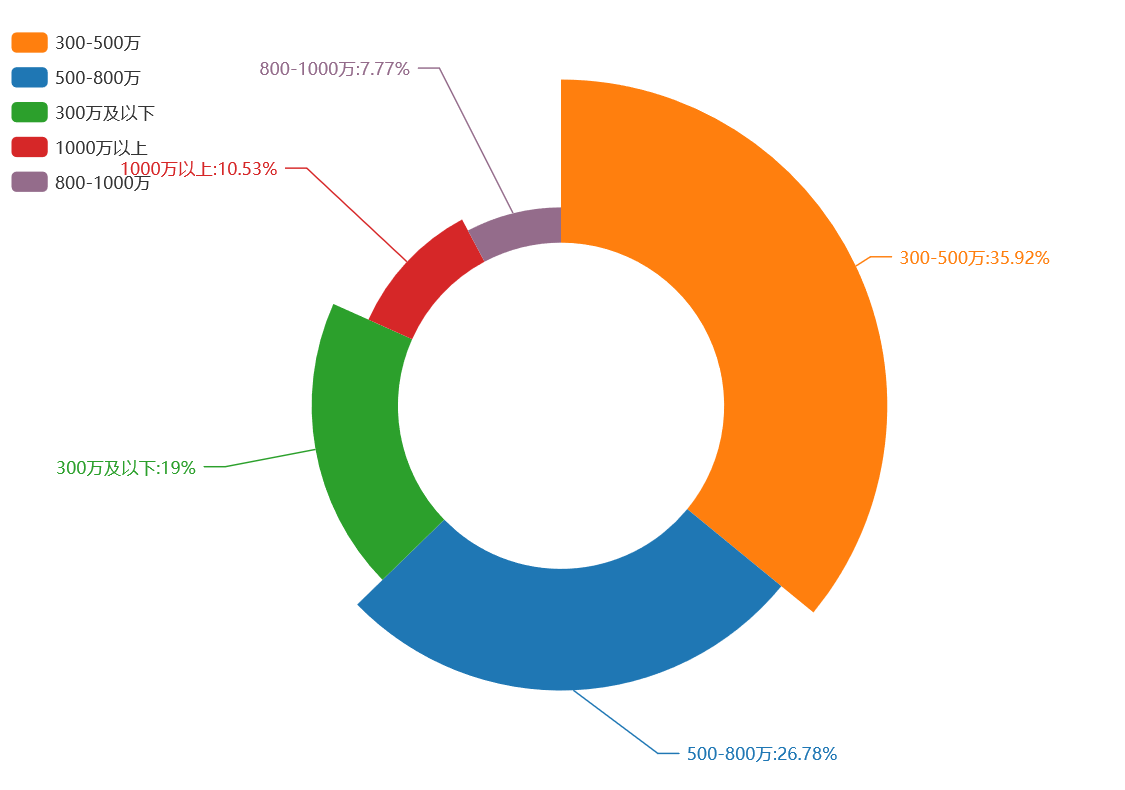
#（3）对总价进行分割（分箱）

#（4）对分箱后的数据进行，分组、统计、排序

#（5）转换成列表对值

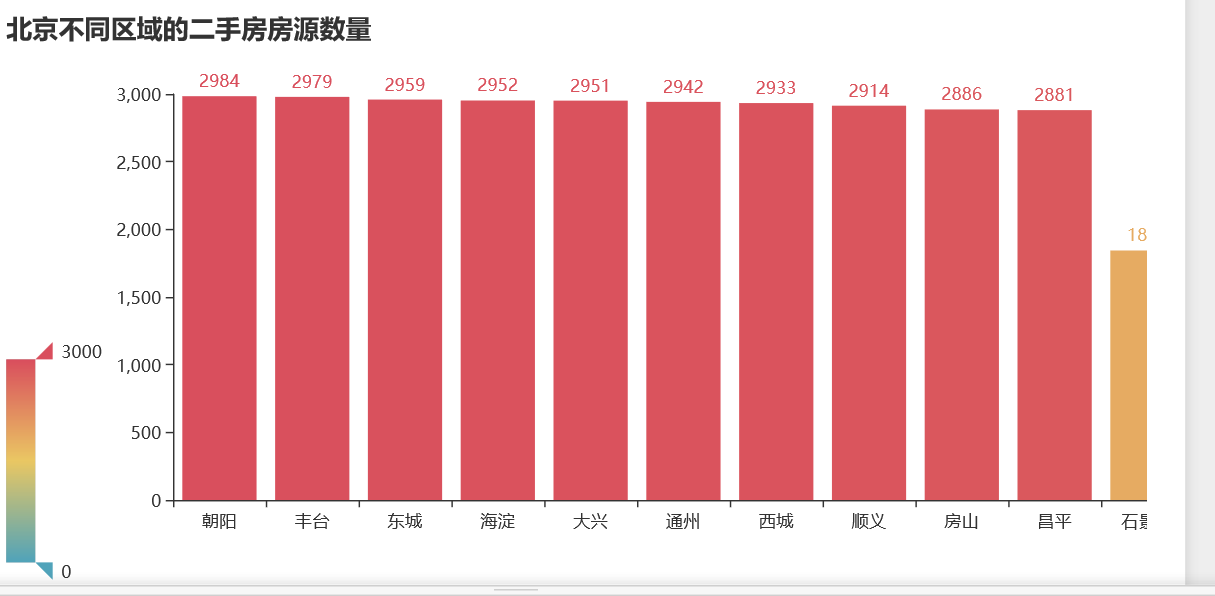
#（6）绘制饼图

#（7）显示图形

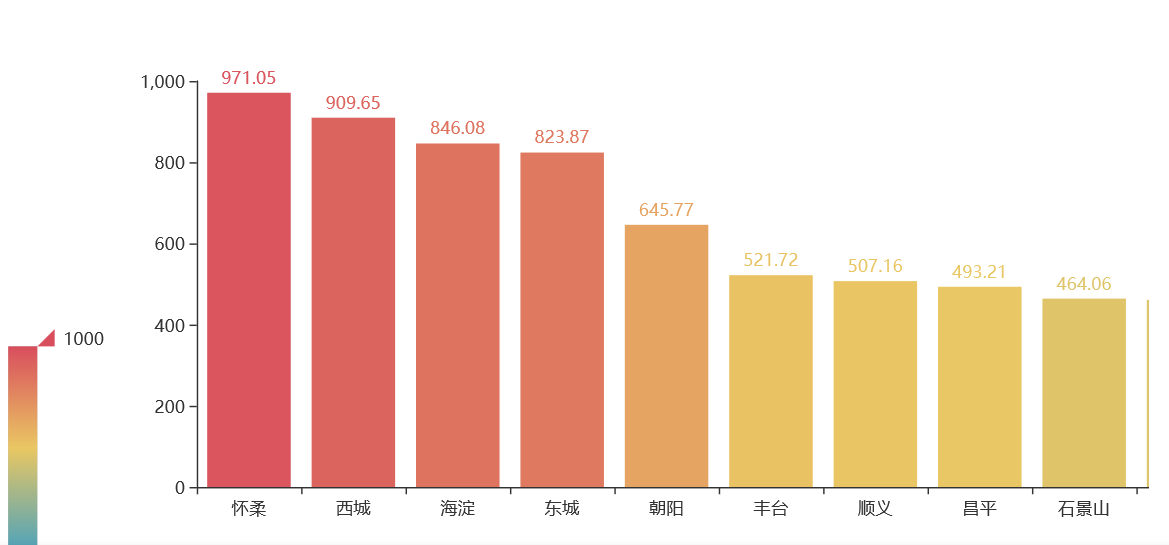


# 5.2 北京不同区域的二手房房源数量

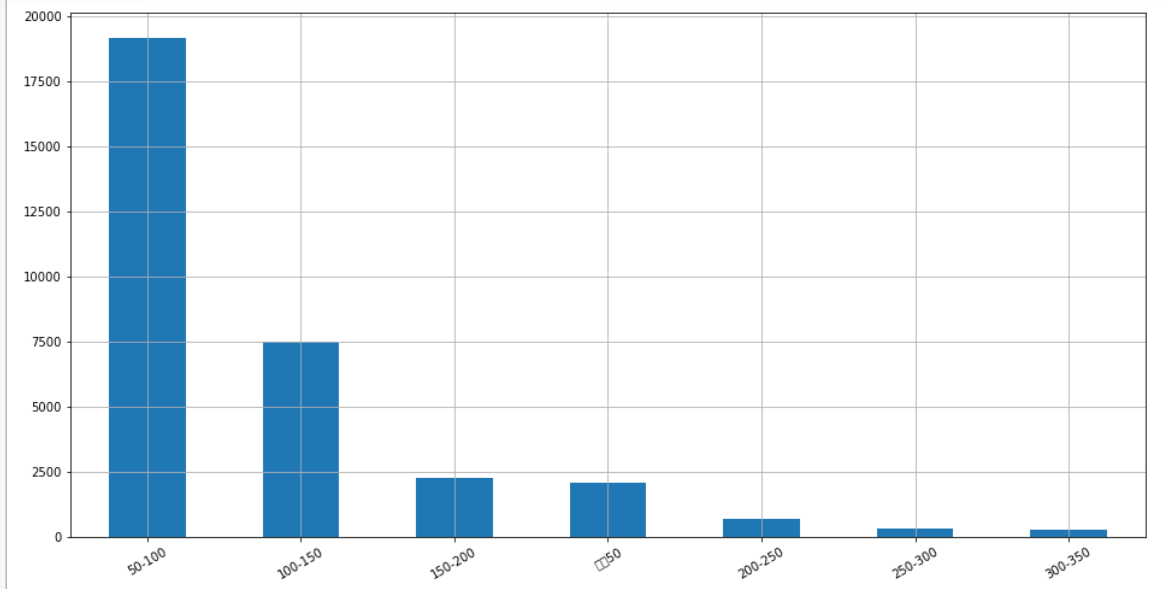
#（1）绘制北京不同区域的二手房房源数量条形图



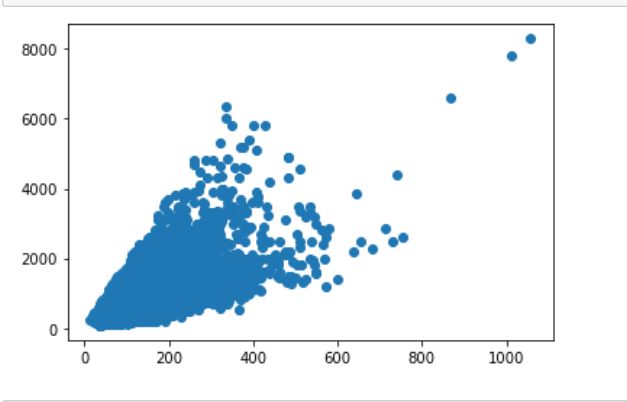
#（2）绘制北京不同区域的二手房均价条形图



#（3）绘制面积分布条形图



#（4）绘制房屋面积和房屋价格的关系散点图

****