

zfx

曾飞雪

2024-10-31

```
library(tidyverse) data <- read.csv("C:/Users/pc/Desktop/2023-09-12_cleaned.csv") data str(data) names(data)
summary(data)
```

```
#1. 发现1大多数房屋都朝向南 # 发现2几乎所有房源的客厅数量都是1-2个 #发现3白沙洲的二手房数量最多
glimpse(data)
```

```
summary(data)
```

```
#直观结论1：房价均值是155万，房价的中位数是137万 #直观结论2：卧室最多的有7间，平均每个房子有2.695个卧室
#直观结论3：武昌的二手房数量比汉口多
```

```
mean(datapricesqm)median(datapricesqm) quantile(datapricesqm)range(datapricesqm) IQR(datapricesqm)var(datapricesqm) sd(datapricesqm)mad(datapricesqm)
```

```
#发现：楼层高低对房屋单价的影响
```

```
#- 发现1 # 大多数房屋的总层数在20层左右。
```

```
#- 发现2 总层数越高，单价也越贵 data<-C data%>% ggplot(aes(x = property_t_height, y = price_sqm))+
geom_point() +
geom_smooth(method = "lm", se = TRUE)
```

```
#发现：2023年小区出现次数前十的名字 name_top10 <- data %>% group_by(property_name) %>%
summarise(count = n()) %>% arrange(desc(count)) %>% head(10) name_top10 ggplot(name_top10, aes(x =
property_name, y = count)) + geom_bar(stat = "identity") + labs(title = "出现最多次数的小区名字", x = "小区名字", y = "出现次数")
#发现1 保利中央公馆和东立国际出现的次数最多。
```

```
#发现2 恒大名都出现的次数较少
```

```
#发现：二手房源中大面积的多还是小面积的房源多
```

```
data <- na.omit(data) databuilding_area <- as.numeric(as.character(data$building_area)) threshold
<- 100
```

```
data$area_category <- ifelse(data$building_area > threshold, "大面积", "小面积") area_counts <-
table(data$area_category) print(area_counts) area_ratio <- prop.table(area_counts)
print(area_ratio) library(ggplot2)
```

```
ggplot(data, aes(x = area_category)) + geom_bar(fill = "red") + labs(title = "房屋面积分类", x = "面积类别", y = "数量")
#- 发现1：小面积的房源比大面积的房源多
```

```
#- 发现2：二者数量相差不大
```

探索问题1

是否近地铁如何影响房屋的单价

```
datanear_subway <- as.factor(datanear_subway == "近地铁")
levels(data$near_subway) <- data("否", "是") region_popularity <- data %>%
group_by(property_region) %>%
summarise(num_properties = n(), .groups = "drop") %>%
arrange(desc(num_properties)) print(region_popularity)
```

```
ggplot(data, aes(x = near_subway, y = price_sqm)) +  
geom_boxplot()  
t_test_result <- t.test(price_sqm ~ near_subway, data = data)  
print(t_test_result) #发现:
```

#- 发现1 是否近地铁影响房屋的单价，靠近地铁的房屋单价要高

#- 发现2 白沙洲近地铁的房源最多

探索问题2

房屋单价是否受房屋面积的影响

```
ggplot(data, aes(x = building_area, y = price_sqm)) +  
geom_point() +  
geom_smooth(method = "lm", se = TRUE) #- 发现1 # 大致上来看房屋面积越大，房屋单价越高  
#- 发现2 # 大多数房源房屋的面积都在100平米左右
```

探索问题3

```
#房屋单价是否受房屋装修影响 ggplot(data, aes(x = decoration, y = price_sqm)) +  
geom_boxplot() #- 发现1 # 目前的房源中，精装房的数量最多  
#- 发现2 # 房屋单价受装修的影响，精装房的房屋单价均值相对而言最高。
```

发现总结

#用1-3段话总结你的发现。

#近几年武汉的房源供应情况以白沙洲，盘龙城等较为偏远的地方居多，市中心房源供应的情况较少，由此可见市区土地有限，人口在逐渐往郊区外扩。#房屋的单价与房屋所在的地段，附近是否有地铁，房屋是否为精装修，房屋的层高以及房屋的面积均相关。一般在其他条件相同的情况下地段越中心，附近有地铁，房屋为精装修，楼高偏高，面积大的房屋单价越高。#虽然房屋的单价受房屋面积的影响，面积打的房屋相对单价越高，但是小面积和大面积的市场供需来看，小面积房屋更多。