

硬核！数据分析 福大水电费问题

——作者：枫、陈用饼、斯普特尼克107号

编辑于2024/10/16

#项目背景

- 福大高价水电费引发关注
- 同学们对水电费的上涨存疑
- 校方并未给出合理的说法或解决方案

#项目结论

- 从整体上看，大约有60%~80%之间的宿舍9月份水电费相比去年9月份增长了，这些宿舍宿舍水电费大约增长了20%，但是还有一些正常的宿舍和减少的宿舍，平均下来，整体平均值大约增长了10%
- 如果你关心的是整体电费开支的变化，推荐参考所有宿舍电费均值的上涨百分比[9.24%（3874间样本）或11.4%（1200间样本）]；如果你关心的是每个宿舍的独立涨幅情况，推荐参考每个宿舍上涨百分比的均值（3874间样本）（20.60%）

#项目目的

- 探究水电费与去年同期相比的上涨幅度
- 探究水电费6-9月月度变化规律

#数据获取

- 作者通过编写脚本，从学生水电费缴费网站——“悦享生活”（公开查询的网站）上获取宿舍水电的历史账单
- 将获取到的账单保存至Excel做数据分析

#实现路径

- 编写Python脚本获取相关数据，源代码如下：

```
1  import time
2  from openpyxl import Workbook
3  import pandas as pd
4  from selenium import webdriver
5  from selenium.webdriver import ChromeOptions
6  from selenium.webdriver.common.by import By
7  from selenium.webdriver.support import expected_conditions as EC
8  from selenium.webdriver.support.ui import WebDriverWait
9
10
11 # 设置 ChromeDriver 的路径
12 def fee(driver, b_num, d_num):
13     driver.get(
14         'https://cloud.life.ccb.com/index_u.jhtml?param=4C4A647F371B43F11DBA91BA7799EE58C59D79B9E8
15     )
16
17     dorm_element = driver.find_element(By.XPATH, "//*[@id='inputTab']/tbody/tr[1]/td[2]/span[2]")
18     dorm_element.click()
19
20     # 选择楼栋
21     # 等待<ul>元素加载完成
```

```
20     # 选择楼栋
21     # 等待<ul>元素加载完成
22     wait = WebDriverWait(driver, timeout=10)
23     ul_element = wait.until(EC.presence_of_element_located((By.ID, "letter#")))
24
25     # 定义要查找的楼号
26     target_building = str(b_num) + "号楼"
27
28     # 查找所有<li>元素
29     li_elements = ul_element.find_elements(By.CLASS_NAME, "letter_content")
30
31     # 遍历<li>元素，并点击匹配的楼号
32     for li in li_elements:
33         if li.text == target_building:
34             driver.execute_script("arguments[0].scrollIntoView(true);", li)
35             li.click()
36             # print(f"已点击{target_building}")
37             break
38     else:
39         print(f"{target_building} 不存在")
```

#实现路径

- 编写Python脚本获取相关数据，源代码如下：

```
41 # 定位到输入框并清空
42 ele = driver.find_element(By.ID, '0000333441')
43 ele.clear() # 清空输入框
44 ele.send_keys(d_num) # 输入新的宿舍号
45
46 # 点击查询按钮
47 dorm_element = driver.find_element(By.XPATH, "//*[@id='chaxunbtn']")
48 dorm_element.click()
49 time.sleep(3)
50 # 点击确认按钮
51 try:
52     confirm_button = WebDriverWait(driver, timeout: 1).until(
53         EC.element_to_be_clickable((By.XPATH, '//div[contains(@class, "singleBtn") and text()="确定"]'))
54     )
55     confirm_button.click()
56 except Exception as e:
57     pass
58
```

```
59 # 直接去表2保存所有数据
60 try:
61     driver.find_element(By.ID, "tab2")
62     lishidingdan = driver.find_element(By.XPATH, "/html/body/div[1]/div[3]/div/span[2]")
63     lishidingdan.click()
64     # 定位<ul>元素
65     ul_element = driver.find_element(By.ID, "tab2")
66     # 定位<ul>内的所有<li>元素
67     li_elements = ul_element.find_elements(By.TAG_NAME, "li")
68     # 存储所有数据的列表
69     data_list = []
70
71     # 遍历所有<li>元素
72     for li in li_elements:
73         # 提取日期信息
74         date = li.find_element(By.CLASS_NAME, "font1").text
75
76         # 提取金额信息
77         amount = li.find_element(By.CSS_SELECTOR, "input.font3").get_attribute("value")
78
```

#实现路径

- 编写Python脚本获取相关数据，源代码如下：

```
79         # 存储数据到字典
80         each_data = {"b_num": b_num, "d_num": d_num, "date": date, "amount": amount}
81         data_list.append(each_data)
82
83
84         return data_list
85     except:
86         return [{"b_num": b_num, "d_num": d_num, "date": "未查询到该宿舍数据",
87                 "amount": "未查询到该宿舍数据"}]
88
89 1个用法
90 def run_one_building(b_num):
91     #新建一个xlsx文档，存在的话覆盖
92     # 创建一个工作簿对象
93     workbook = Workbook()
94
95     # 获取默认的工作表
96     sheet = workbook.active
97     sheet.title = "Sheet1"
```

```
98     # 保存工作簿
99     file_name = '{}号楼.xlsx'.format(b_num)
100     workbook.save(file_name)
101
102     # 初始化一次浏览器
103     option = ChromeOptions()
104     option.add_experimental_option(name='excludeSwitches', value=['enable-automation'])
105     driver = webdriver.Chrome(chrome_options=option)
106     d_num_list = []
107     for i in range(1, 7):
108         for j in ['01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12',
109                 '18', '19', '20']:
110             d_num_list.append(int(str(i) + j))
111     for d_num in d_num_list:
112         data = fee(driver, b_num, d_num)
113         if data is not None:
114             pd_data = pd.DataFrame(data)
115             file_path = file_name
```


#实现路径

- 编写Python脚本获取相关数据，源代码如下：

```
116         # 使用 ExcelWriter 追加数据
117         with pd.ExcelWriter(file_path, mode='a', engine='openpyxl', if_sheet_exists='overlay') as writer:
118             # 假设工作表名称为 'Sheet1'
119             # 获取最大行数，找到空行位置以追加数据
120             pd_data.to_excel(writer, sheet_name='Sheet1', index=False, header=False,
121                               startrow=writer.sheets['Sheet1'].max_row)
122
123         driver.quit()
124
125
126  ▶ if __name__ == '__main__':
127         building_left = [48, 49,
128                           5, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59,
129                           6, 60, 61, 62, 63, 7, 8, 9]
130         for each_num in building_left:
131             run_one_building(each_num)
```

#原始数据

- 运行完脚本，我们就得到了各个楼栋宿舍的水电费数据

名称	修改日期	类型	大小	F1204					
3号楼.xlsx	2024/10/16...	XLSX 工作表	33 KB		A	B	C	D	E
4号楼.xlsx	2024/10/16...	XLSX 工作表	34 KB	1142	1	613	2023年11月	¥89.01	
10号楼.xlsx	2024/10/16...	XLSX 工作表	41 KB	1143	1	613	2023年10月	¥146.04	
11号楼.xlsx	2024/10/16...	XLSX 工作表	33 KB	1144	1	613	2023年9月	¥479.70	
12号楼.xlsx	2024/10/16...	XLSX 工作表	27 KB	1145	1	613	2023年8月	¥331.53	
13号楼.xlsx	2024/10/16...	XLSX 工作表	30 KB	1146	1	614	2024年9月	¥599.57	
14号楼.xlsx	2024/10/16...	XLSX 工作表	34 KB	1147	1	614	2024年8月	¥337.75	
15号楼.xlsx	2024/10/16...	XLSX 工作表	31 KB	1148	1	614	2024年7月	¥419.47	
16号楼.xlsx	2024/10/16...	XLSX 工作表	28 KB	1149	1	614	2024年6月	¥369.37	
20号楼.xlsx	2024/10/16...	XLSX 工作表	31 KB	1150	1	614	2024年5月	¥139.65	
21号楼.xlsx	2024/10/16...	XLSX 工作表	31 KB	1151	1	614	2024年4月	¥161.50	
22号楼.xlsx	2024/10/16...	XLSX 工作表	32 KB	1152	1	614	2024年3月	¥50.10	
23号楼.xlsx	2024/10/16...	XLSX 工作表	29 KB	1153	1	614	2024年1-2月	¥107.67	
24号楼.xlsx	2024/10/16...	XLSX 工作表	31 KB	1154	1	614	2023年12月	¥198.76	
25号楼.xlsx	2024/10/16...	XLSX 工作表	29 KB	1155	1	614	2023年11月	¥231.34	
26号楼.xlsx	2024/10/16...	XLSX 工作表	27 KB	1156	1	614	2023年10月	¥273.57	
27号楼.xlsx	2024/10/16...	XLSX 工作表	24 KB	1157	1	614	2023年9月	¥610.31	
28号楼.xlsx	2024/10/16...	XLSX 工作表	24 KB	1158	1	614	2023年8月	¥20.79	
29号楼.xlsx	2024/10/16...	XLSX 工作表	31 KB	1159	1	615	2024年7月	¥87.41	
30号楼.xlsx	2024/10/16...	XLSX 工作表	28 KB	1160	1	615	2024年6月	¥275.56	
31号楼.xlsx	2024/10/16...	XLSX 工作表	30 KB	1161	1	615	2024年5月	¥120.99	
				1162	1	615	2024年4月	¥109.27	

#数据整理

- 使用Excel与Python将同时有去年9月份与今年6-9月份水电费数据的宿舍筛选出来，一共1200间宿舍，作为此次项目的研究对象

	A	B	C	D	E
1	宿舍	缴费月份	缴费金额		
2	1#101	2024年9月	¥533. 53		
3		2024年8月	¥223. 33		
4		2024年7月	¥273. 43		
5		2024年6月	¥356. 92		
6		2023年9月	¥534. 60		
7	1#104	2024年9月	¥407. 75		
8		2024年8月	¥261. 70		
9		2024年7月	¥294. 22		
10		2024年6月	¥280. 36		
11		2023年9月	¥346. 59		
12	1#106	2024年9月	¥604. 03		
13		2024年8月	¥309. 67		
14		2024年7月	¥437. 06		
15		2024年6月	¥346. 45		
16		2023年9月	¥507. 64		
17		2024年9月	¥535. 47		
18		2024年8月	¥79. 42		

5983		2024年8月	¥194. 29		
5984	47#602	2024年7月	¥353. 05		
5985		2024年6月	¥184. 16		
5986		2023年9月	¥134. 25		
5987	47#604	2024年9月	¥307. 42		
5988		2024年8月	¥208. 01		
5989		2024年7月	¥319. 15		
5990	47#605	2024年6月	¥183. 43		
5991		2023年9月	¥181. 83		
5992		2024年9月	¥97. 01		
5993	47#606	2024年8月	¥90. 61		
5994		2024年7月	¥133. 25		
5995		2024年6月	¥81. 55		
5996	47#606	2023年9月	¥109. 27		
5997		2024年9月	¥180. 30		
5998		2024年8月	¥102. 21		
5999	47#606	2024年7月	¥197. 48		
6000		2024年6月	¥128. 00		
6001		2023年9月	¥100. 95		
6002					

#数据分析

- 分析去年9月与今年9月的水电费上涨幅度

上涨11.4%

去年
9
月

今年
9
月

5952	47#503	2024年9月	¥120.07
5957	47#504	2024年9月	¥218.61
5962	47#505	2024年9月	¥56.84
5967	47#507	2024年9月	¥75.69
5972	47#508	2024年9月	¥173.23
5977	47#601	2024年9月	¥110.33
5982	47#602	2024年9月	¥324.21
5987	47#604	2024年9月	¥307.42
5992	47#605	2024年9月	¥97.01
5997	47#606	2024年9月	¥180.30

5951	47#502	2023年9月	¥151.91
5956	47#503	2023年9月	¥90.48
5961	47#504	2023年9月	¥129.79
5966	47#505	2023年9月	¥72.31
5971	47#507	2023年9月	¥216.20
5976	47#508	2023年9月	¥235.46
5981	47#601	2023年9月	¥132.52
5986	47#602	2023年9月	¥134.25
5991	47#604	2023年9月	¥181.83
5996	47#605	2023年9月	¥109.27
6001	47#606	2023年9月	¥100.95

1047376			
1047377			
1047378			
1047379			

在 1047374 个记录中筛选出 1200 个 平均值=288.2616 计数=1200 求和=34万5913.92

1047376			
1047377			
1047378			
1047379			

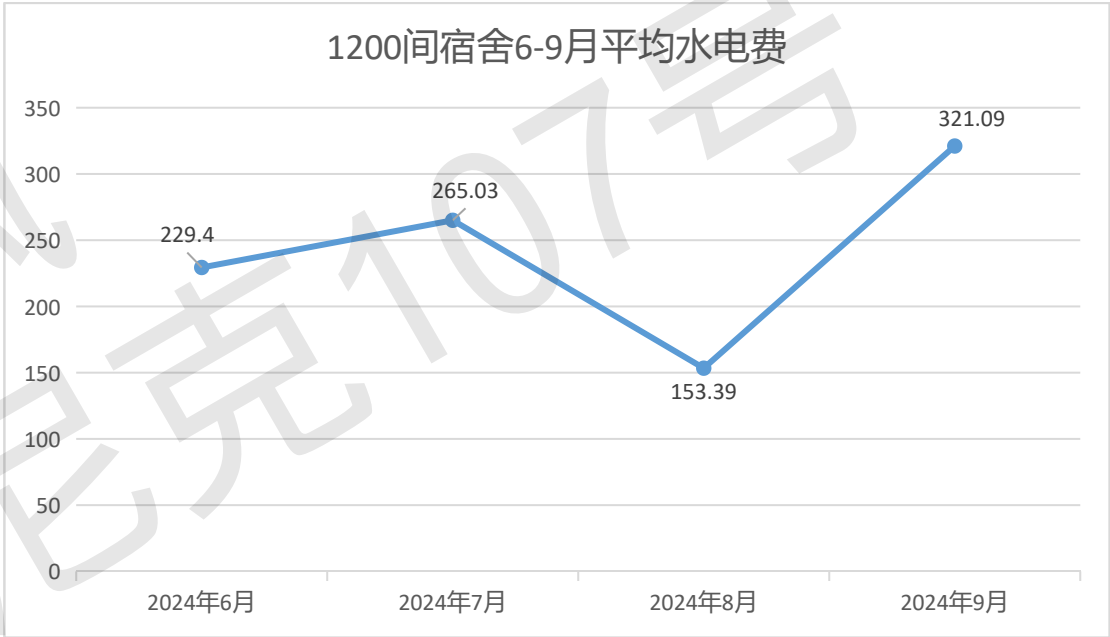
在 1047374 个记录中筛选出 1200 个 平均值=321.086033333 计数=1200 求和=38万5303.24

#数据分析

- 各个宿舍今年月度变化情况

9月水电较6-8月份均有所上涨
两个注意点：

- ①7-8月份为学生暑假，宿舍用电应该偏少。
 - ②9月计费周期为8.24-9.29，共36天比较长
- 综上，目前很难下定结论今年9月份的电费一定存在异常。



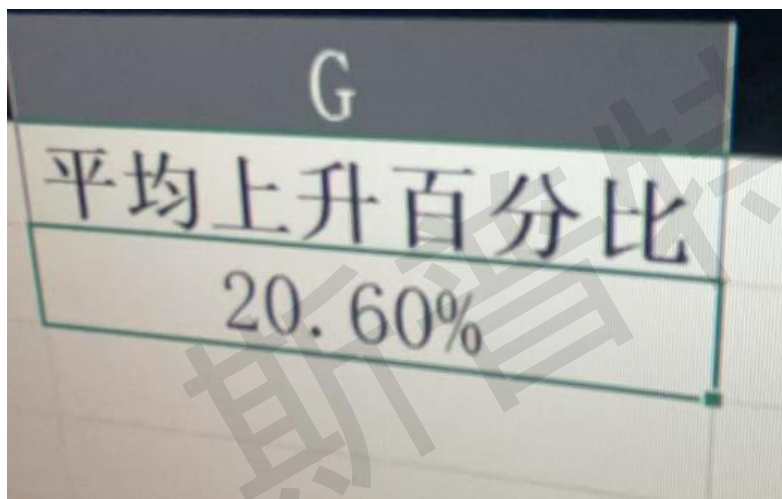
1200间宿舍6-9月平均水电费			
2024年6月	2024年7月	2024年8月	2024年9月
229.4	265.03	153.39	321.09

#总结

1. 目前很难下定结论今年9月份的电费一定存在异常
 2. 所选1200间宿舍（样本）去年9月与今年9月的水电费上涨幅度约为11.4%
 3. 考虑到一些因素，所选1200间宿舍（样本）今年6-9月水电费变化幅度在合理范围内
- 声明：
 - 数据由个人通过公开渠道获取
 - 数据、代码可供交流学习
 - 若存在不足或者错误，望指正
 - 以上均为个人观点仅供参考（枫）

#其他补充信息汇总

- 有一份样本量更大（这份样本有3874间宿舍，而作者只选取同时有6-9月份水电费账单的宿舍，故样本量只有1200间）的数据显示，若单独看去年9月份的数据与今年九月份的数据，水电费上涨幅度约为**20.60%**（左图），该样本还统计了各区涨幅的差异（右图）



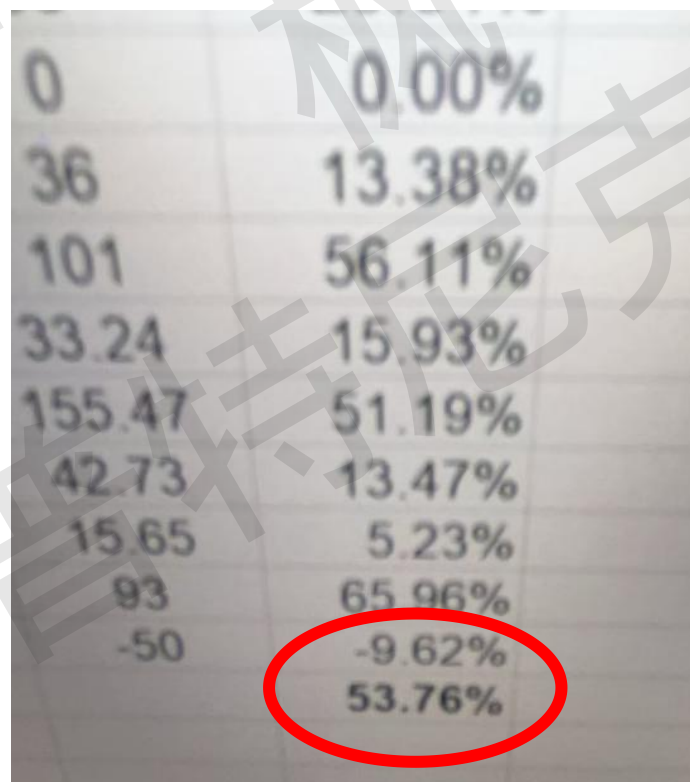
A photograph of a digital display showing a table of electricity and water fee percentage increases by district. The table has two columns: '区号' (District Number) and '水电费上升百分比' (Percentage increase in electricity and water fees). The data is as follows:

区号	水电费上升百分比
1区	14.89%
2区	20.32%
3区	24.06%
4区	21.75%
5区	27.84%

数据有出入，供参考，欢迎大家积极讨论.....

#其他补充信息汇总

- 同学做的问卷数据显示，对比去年9月，宿舍水电的平均上涨幅度为53.76%

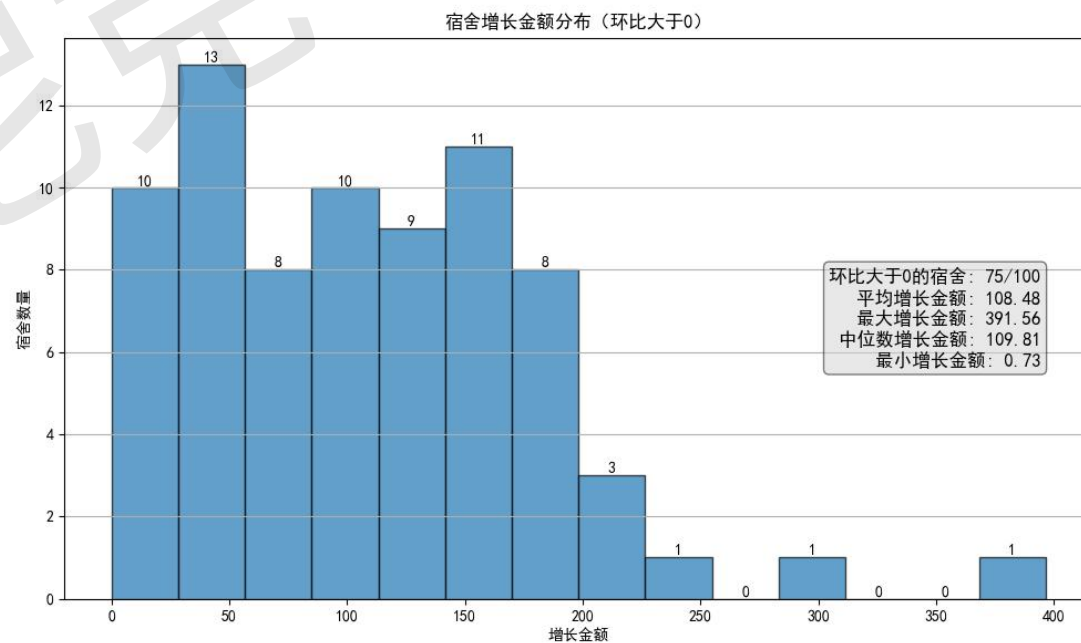
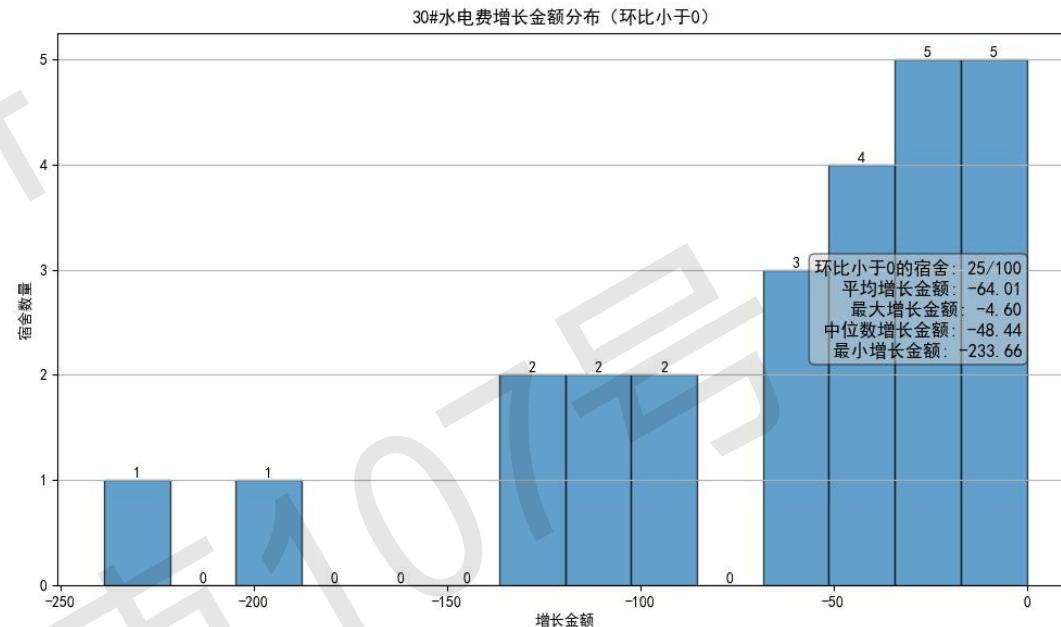
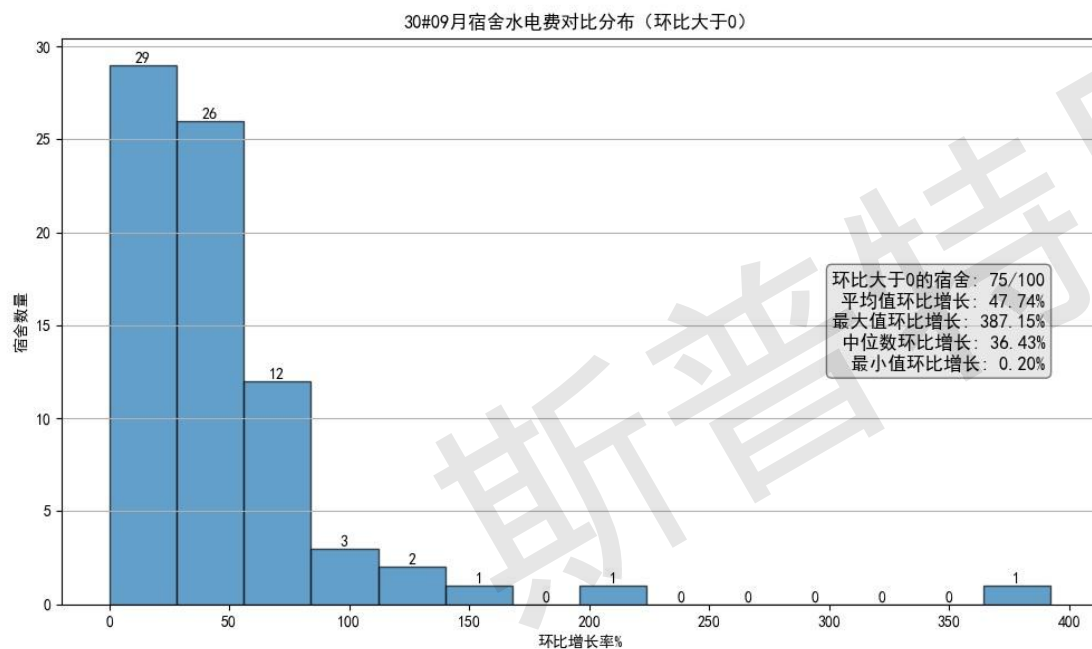


0	0.00%
36	13.38%
101	56.11%
33.24	15.93%
155.47	51.19%
42.73	13.47%
15.65	5.23%
93	65.96%
-50	-9.62%
	53.76%

数据有出入，供参考，欢迎大家积极讨论.....

#其他补充信息汇总

- 以30#为样本，只统计去年与今年9月份的数据结果

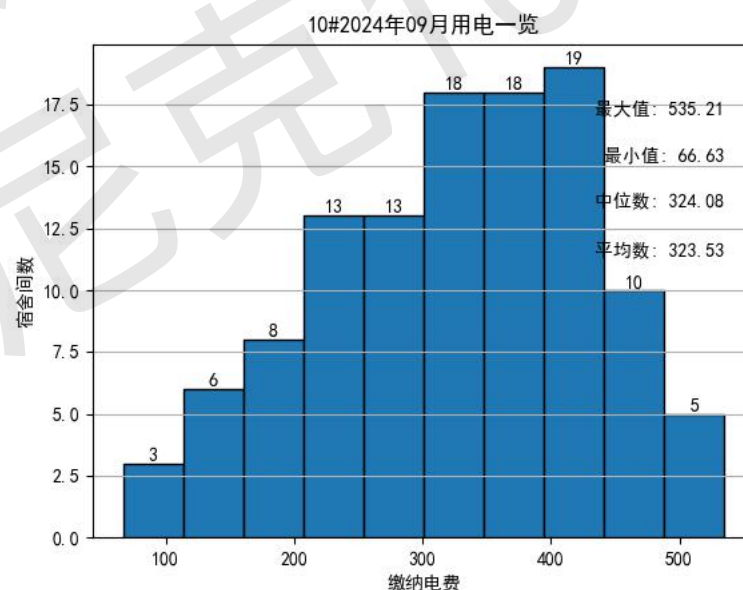
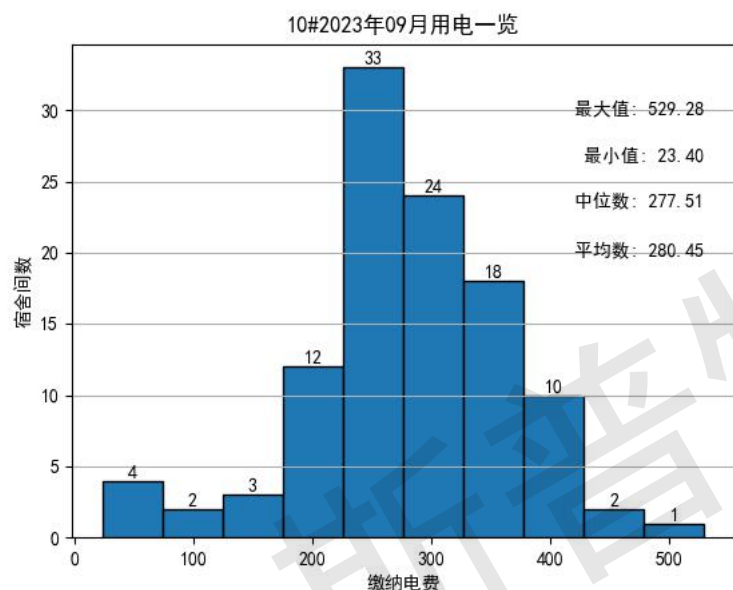


数据有出入，供参考，欢迎大家积极讨论.....

#其他补充信息汇总

平均值上涨15.4%

- 以10#为样本，只统计去年与今年9月份的数据结果

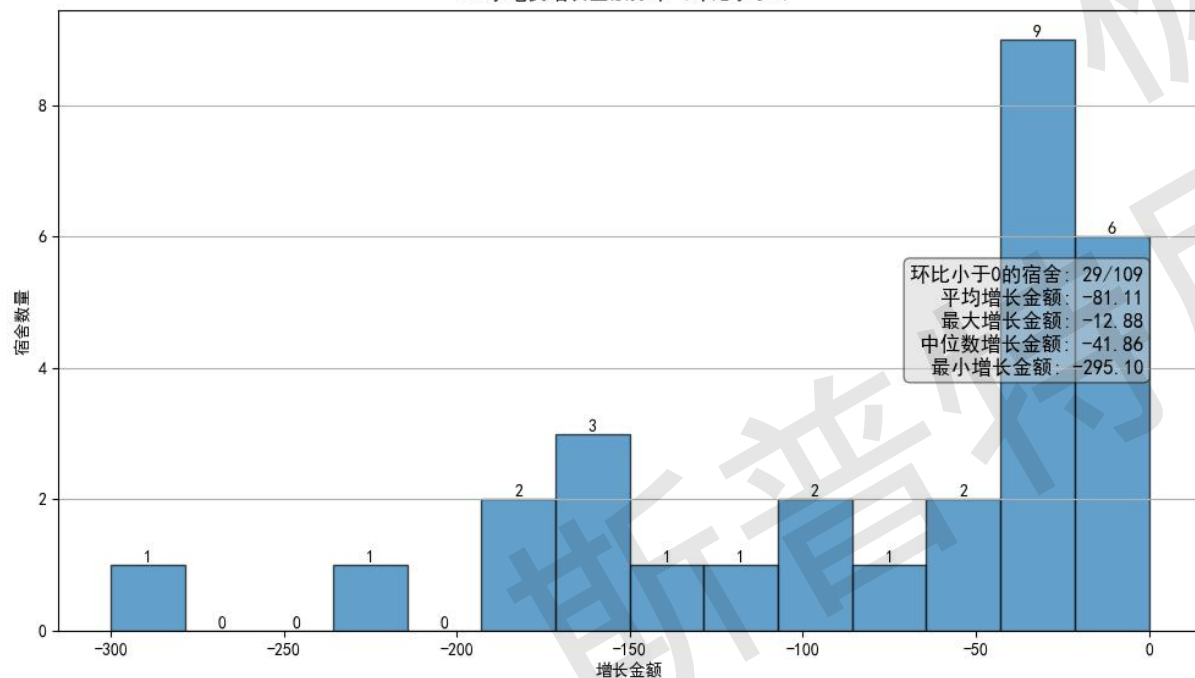


数据有出入，供参考，欢迎大家积极讨论.....

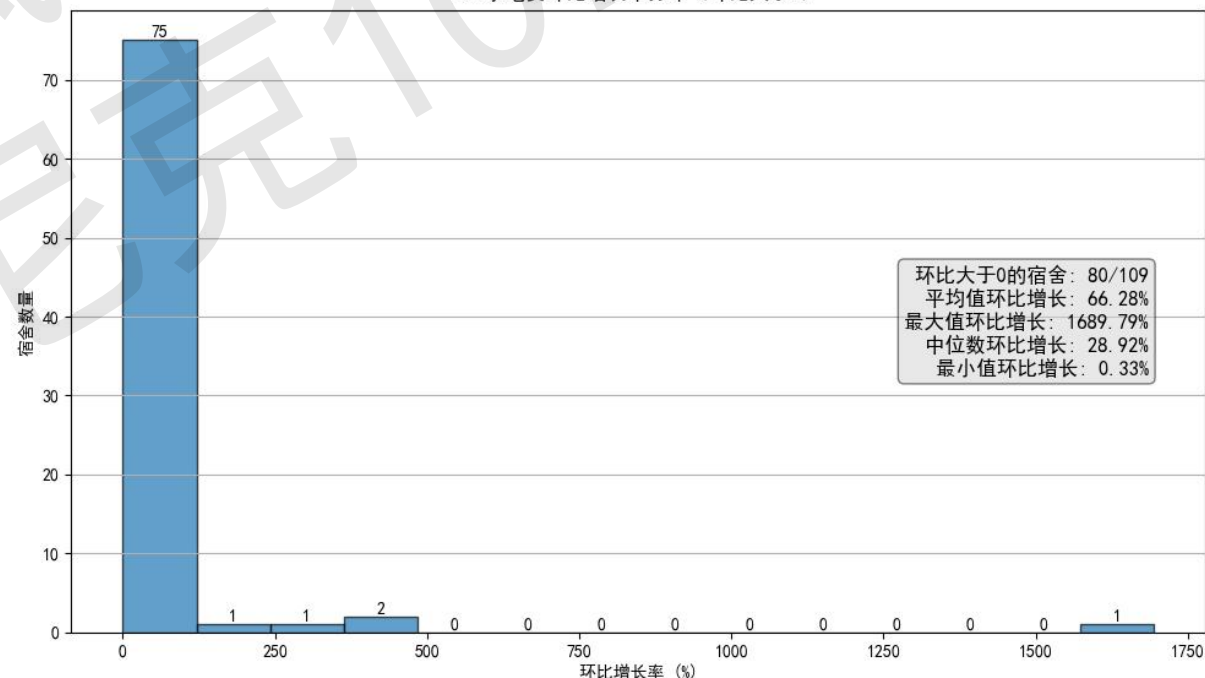
#其他补充信息汇总

- 以10#为样本，只统计去年与今年9月份的数据结果

10#水电费增长金额分布（环比小于0）



10#水电费环比增长率分布（环比大于0）



数据有出入，供参考，欢迎大家积极讨论.....

#项目结论

- 从整体上看，大约有60%~80%之间的宿舍9月份水电费相比去年9月份增长了，这些宿舍宿舍水电费大约增长了20%，但是还有一些正常的宿舍和减少的宿舍，平均下来，整体平均值大约增长了10%
- 如果你关心的是整体电费开支的变化，推荐参考所有宿舍电费均值的上涨百分比[9.24%（3874间样本）或11.4%（1200间样本）]；如果你关心的是每个宿舍的独立涨幅情况，推荐参考每个宿舍上涨百分比的均值（3874间样本）（20.60%）