

数据分析业务基础

走进数据分析的世界

- 为什么用Python进行数据分析
- 如何成为数据分析师
- 数据分析师的必备技能

个人简介

Heywhale 和鲸



LeeBer老师

某在线教育平台数据分析教研负责人,

国内某房地产开发商数据分析负责人.

中国农业银行Python开发工程师

各大银行企业数据分析实训讲师

丰富的从0-1业务数据分析经验, 擅长数据分析技术与业务结合. 讲课风格通俗易懂, 风趣幽默

为什么用Python进行数据分析

为什么用python做数据分析？

Heywhale 和鲸

如果你对
Excel、Power
BI 但它们的
功能相对局限单一。
怎么解决



比如Excel、Tableau
功能相对局限单一。

为什么用python做数据分析?

简单, 易学, 适合初学者

为什么用python做数据分析?

```
1 using System;
```

```
1 class HelloWorld
```

Java

```
print('Hello World')
```

```
6 }
```

```
7 }
```

```
12 }
```

为什么用python做数据分析？

活跃的科学计算社区和强大的第三方库

为什么用python做数据分析?

Heywhale 和鲸

和鲸
HeyWhale.com

kaggle

为什么用python做数据分析?

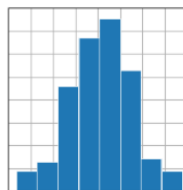
Heywhale 和鲸

matplotlib

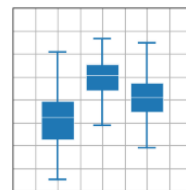
Plot types Examples Tutorials Reference User guide Develop Release notes



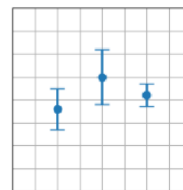
stem(x, y)
step(x, y)
fill_between(x, y1, y2)
imshow(Z)
pcolormesh(X, Y, Z)
contour(X, Y, Z)
contourf(X, Y, Z)
barbs(X, Y, U, V)
quiver(X, Y, U, V)
streamplot(X, Y, U, V)
hist(x)
boxplot(X)
errorbar(x, y, yerr, xerr)
violinplot(D)
eventplot(D)
hist2d(x, y)



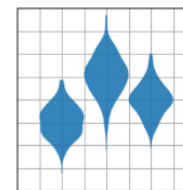
hist(x)



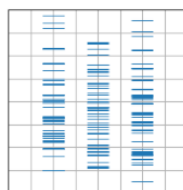
boxplot(X)



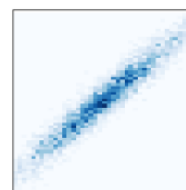
errorbar(x, y, yerr, xerr)



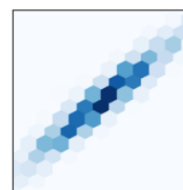
violinplot(D)



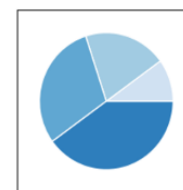
eventplot(D)



hist2d(x, y)



hexbin(x, y, C)



pie(x)

On this page

Basic

Plots of arrays and fields

Statistics plots

Unstructured coordinates

为什么用python做数据分析?

Heywhale 和鲸

boken



pyecharts

matplotlib



Seaborn



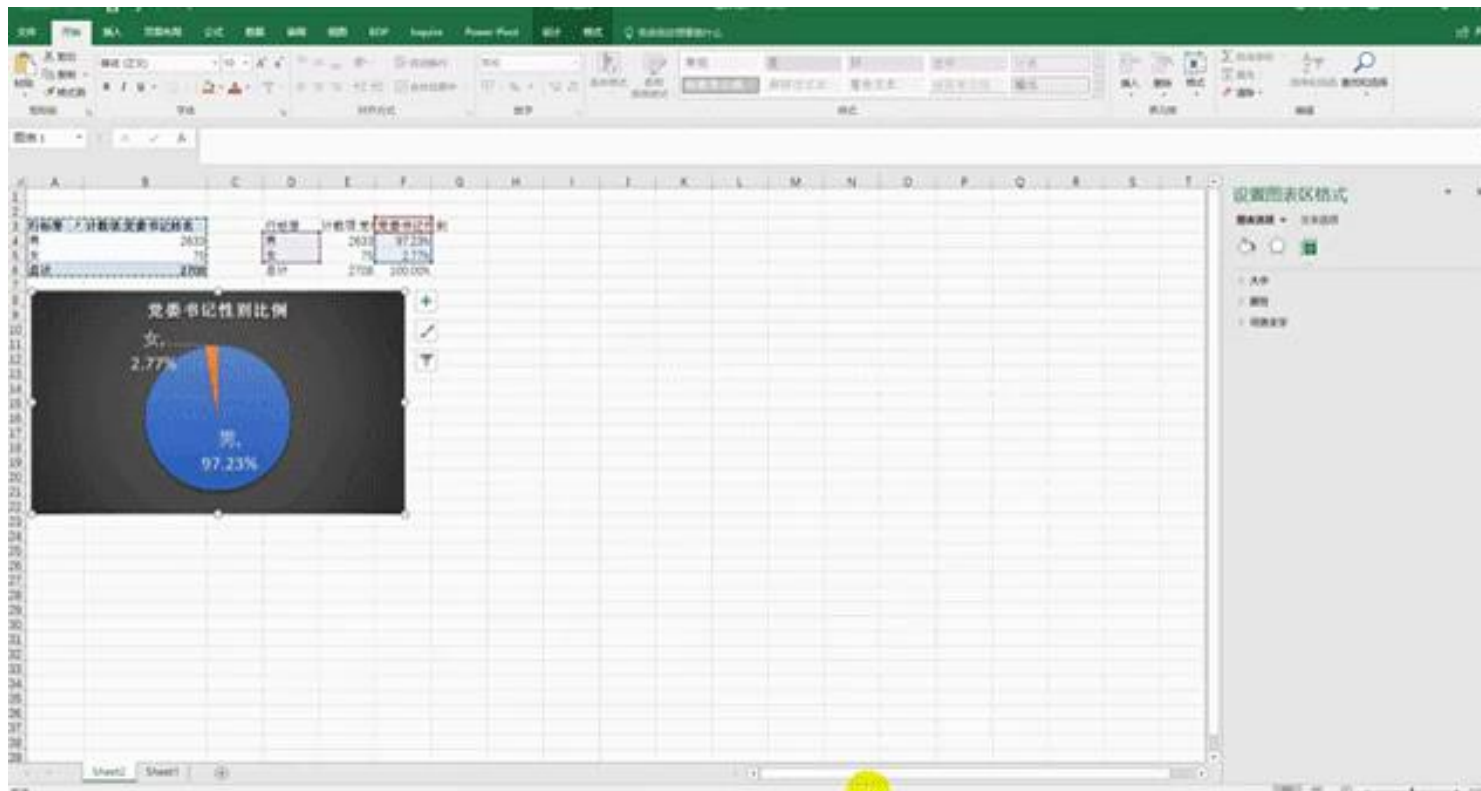
plotly

为什么用python做数据分析？

流程可控，工作高效

为什么用python做数据分析?

Heywhale 和鲸



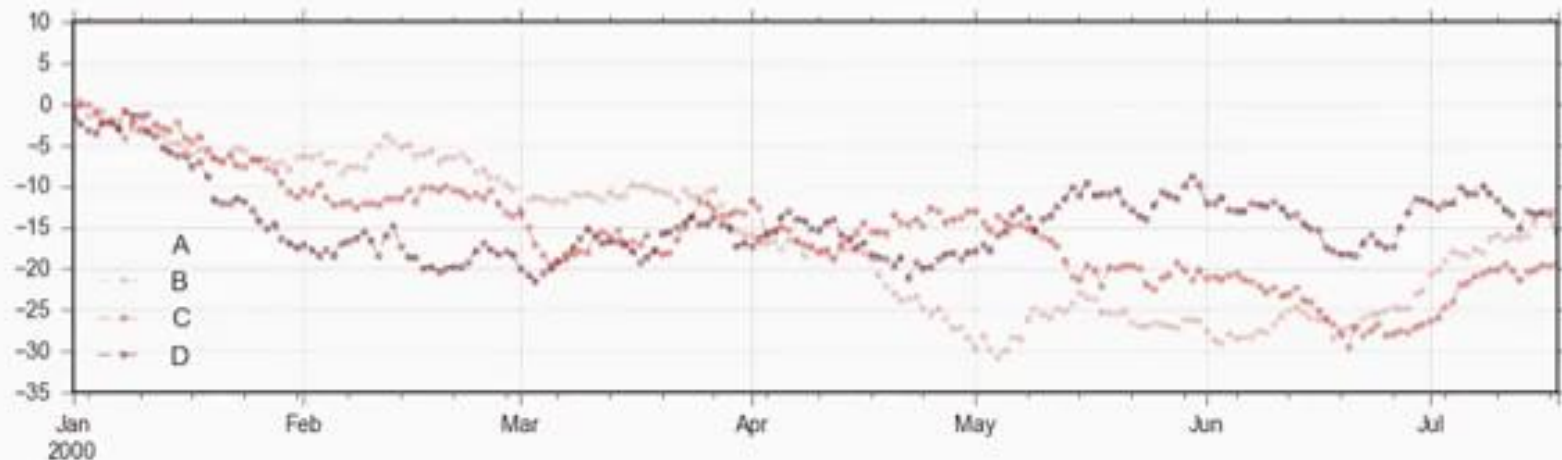
为什么用python做数据分析?

Heywhale 和鲸

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [5]: # 整体风格样式
%matplotlib inline
import matplotlib.style as plt
# 查看样式列表
# ['seaborn-ticks', 'ggplot', 'seaborn-bright', 'seaborn-poster', 'seaborn-dark-palette', 'buh', 'classic', 'seaborn-white', 'fivethirtyeight',
# 'seaborn-paper', 'seaborn-darkgrid', 'seaborn-deep', 'seaborn-talk', 'seaborn-dark', 'seaborn-colorblind', 'seaborn-whitegrid',
# 'grayscale', 'seaborn-muted', 'seaborn-pastel', 'seaborn-notebook', 'dark_background']
# plt.use('ggplot')
# plt.use('seaborn-darkgrid')
# plt.use('dark_background')
plt.use('seaborn-ticks')
df = pd.DataFrame(np.random.randn(200, 4), index=pd.date_range('1/1/2000', periods=200), columns=list('ABCD'))
df = df.cumsum()
df.plot(kind='line', style='--', alpha=0.4, use_index=True, grid=True, figsize=(12, 3), legend=True, subplots=False, colormap='Reds')
```

Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0xb4dc940>



为什么用python做数据分析？

人工智能时代的通用语言

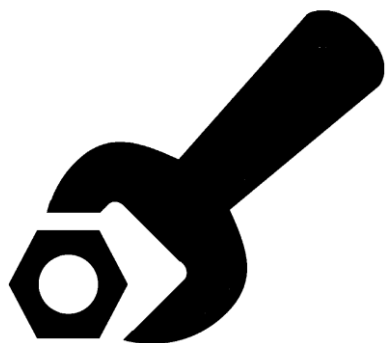
为什么用python做数据分析?

Heywhale 和鲸



如何成为数据分析师

如何成为数据分析师?



工具



业务



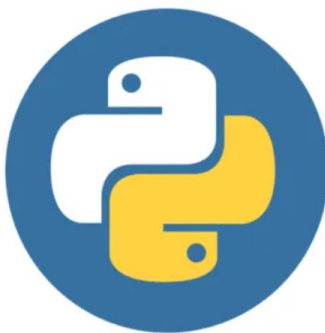
思维

清洗工具

Heywhale 和鲸

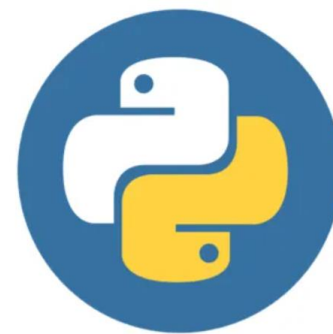


SQL



可视化工具

Heywhale 和鲸



如何理解数据？

常用的指标有哪些？

如何选择指标？

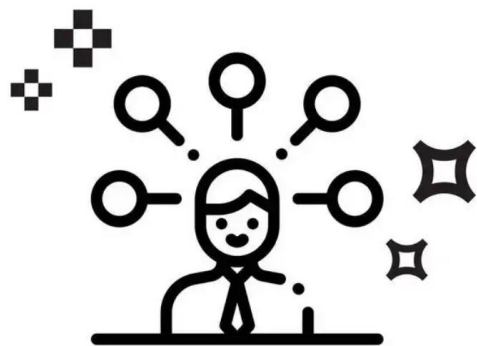
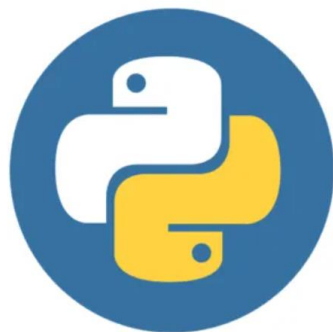
思维

1. RMF分析
2. 漏斗分析
3. 归因分析
4. 留存分析
5. LTV分析
6. 路径分析
7. 趋势分析
8. 热力分析
9.

数据分析师必备技能

数据分析师必备的技能

Heywhale 和鲸



Multi Thinking



结束语

Heywhale 和鲸

感谢大家聆听

