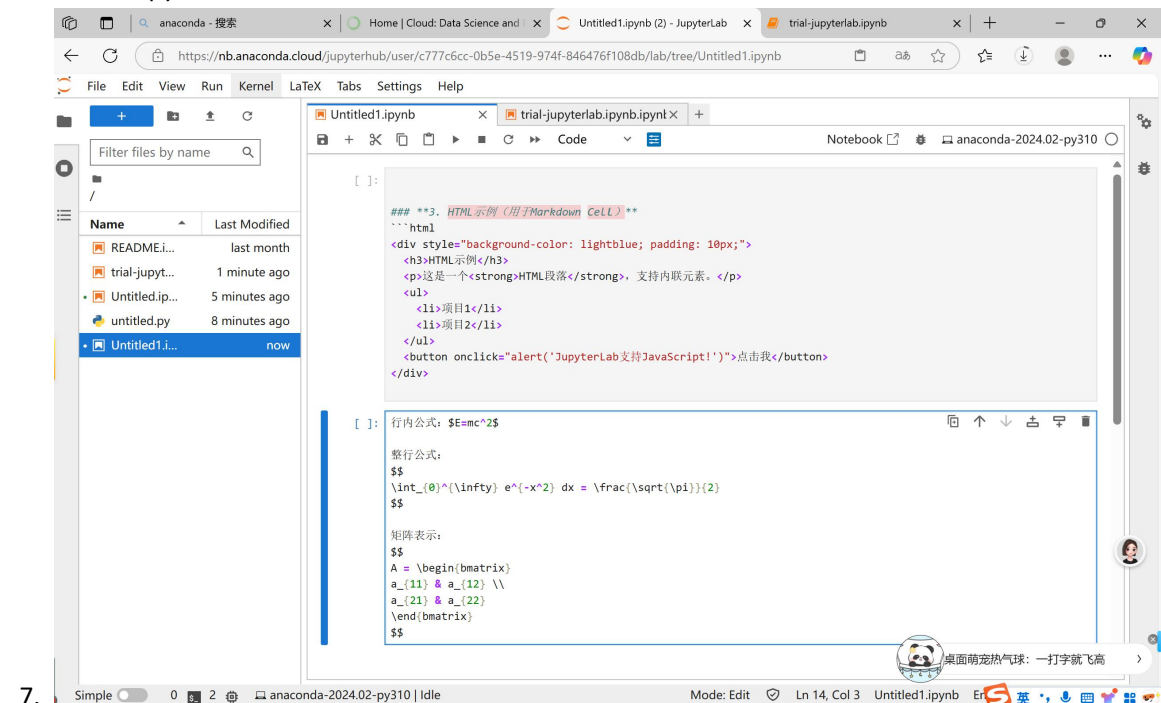


# Week07

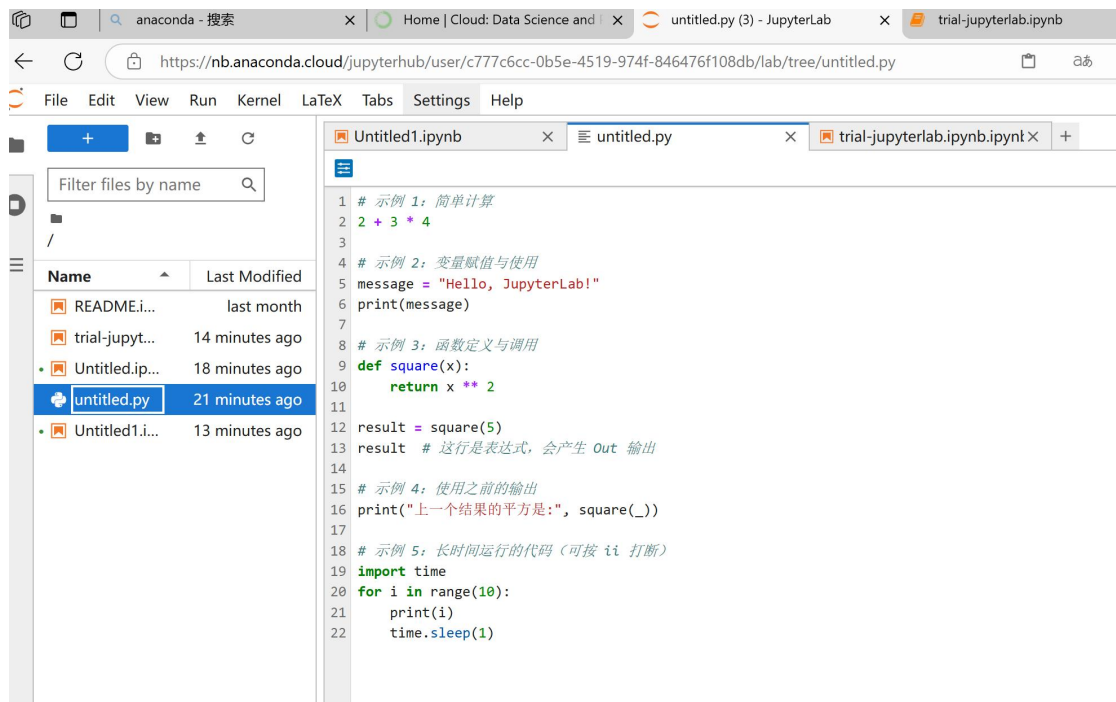
LaTeX 数学公式示例 以下 LaTeX 数学公式代码可以复制到 Markdown 单元格中渲染。

操作步骤总结:

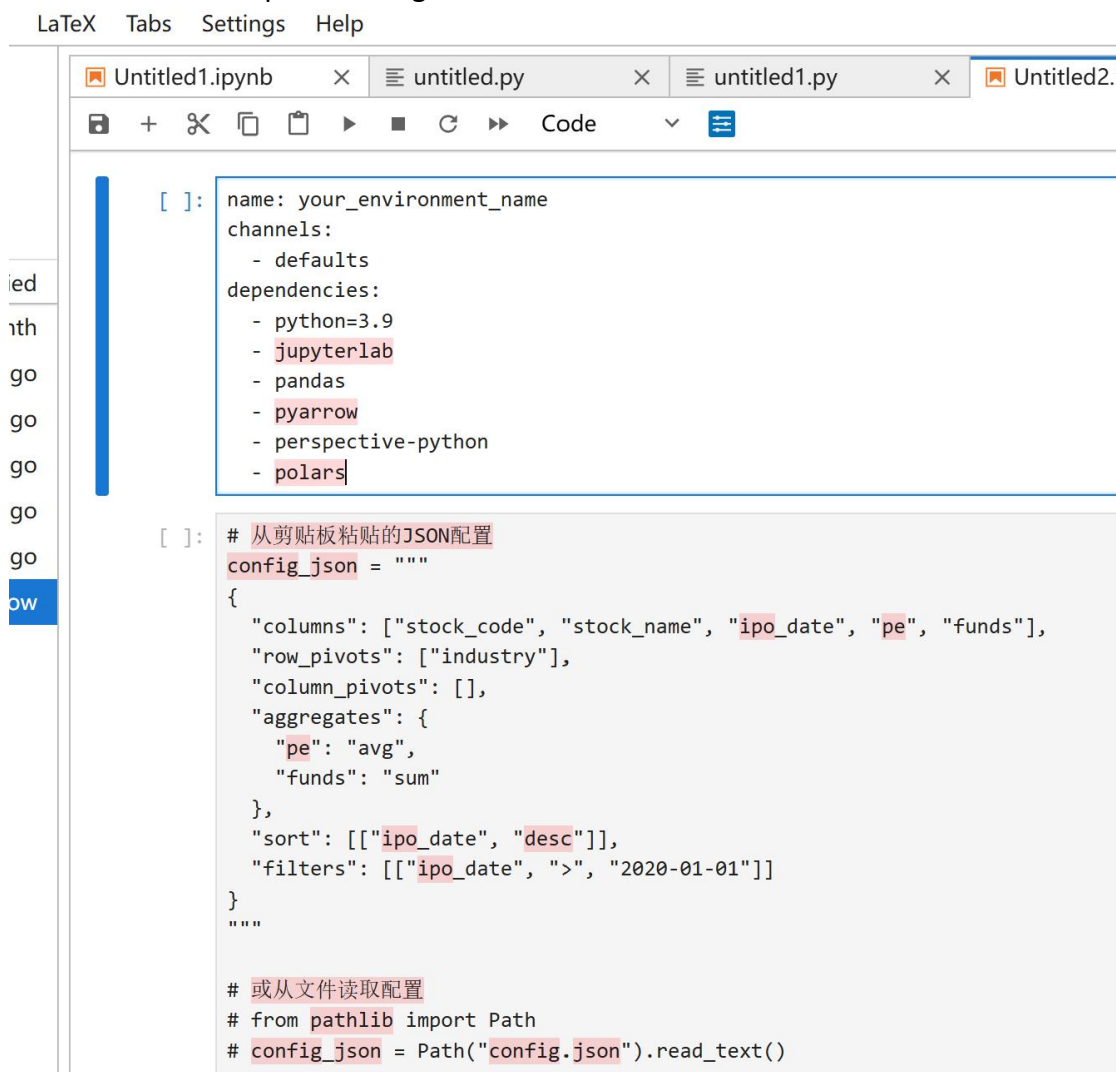
1. 创建并命名 Notebook: 点击 JupyterLab 左侧工具栏的「+」新建 Notebook, 选择 Python 内核点击顶部文件名 (默认为 Untitled.ipynb), 重命名为 trial-jupyterlab.ipynb。
2. 练习 Cell 操作: 编写上述 Python 代码, 按 Shift+Enter 运行并下移, 按 Esc 进入命令模式, 练习 j、k、a、b、dd 等快捷键切换到编写模式 (Enter), 按 Ctrl+Shift+- 切分单元格。
3. 测试 Kernel 和输出: 运行长时间代码 (示例 4), 按 ii 打断
4. 使用 00 重启 Kernel, 清空输出后重新运行所有 Cell。
5. Markdown 和 LaTeX 实践: 创建 Markdown Cell, 粘贴上述 Markdown、HTML、LaTeX 代码, 按 Shift+Enter 渲染。
6. 关闭 JupyterLab



## 1 修改 environment.yml 文件



## 2 数据处理与 PerspectiveWidget 初始化



### 3. 应用导出的配置（示例）

```
Untitled1.ipynb x untitled.py x untitled1.py x Untitled2.ipynb x
[ ]: # 或从文件读取配置
      # from pathlib import Path
      # config_json = Path("config.json").read_text()

      import json
      config = json.loads(config_json)

      # 使用配置初始化PerspectiveWidget
      widget_with_config = PerspectiveWidget(d1.to_pandas(), **config)
      widget_with_config

      [ ]: config_json = """
            {
              "columns": ["stock_code", "stock_name", "ipo_date", "pe", "funds"],
              "row_pivots": ["industry"],
              "column_pivots": [],
              "aggregates": {
                "pe": "avg",
                "funds": "sum"
              },
              "sort": [["ipo_date", "desc"]],
              "filters": [["ipo_date", ">", "2020-01-01"]]
            }
            """

            # 或从文件读取配置
            # from pathlib import Path
            # config_json = Path("config.json").read_text()

            import json
            config = json.loads(config_json)
```

### 4. 数据可视化示例 直方图（Histogram）

```
Untitled1.ipynb x untitled.py x untitled1.py x Untitled2.ipynb
Code
# 使用配置初始化PerspectiveWidget
widget_with_config = PerspectiveWidget(d1.to_pandas(), **config)
widget_with_config

[ ]: # 创建分桶列并生成直方图
d1_hist = d1.with_columns(
    pl.col("pe").bin(10).alias("pe_bucket") # 将PE分为10个桶
)

hist_config = {
    "columns": ["pe_bucket", "stock_code"],
    "row_pivots": ["pe_bucket"],
    "aggregates": {"stock_code": "count"},
    "plugin": "y_bar" # Y Bar视图
}

PerspectiveWidget(d1_hist.to_pandas(), **hist_config)

[ ]: # 按月汇总融资额和市盈率
d1_time = d1.with_columns(
    pl.col("ipo_date").dt.strftime("%Y-%m").alias("month")
).groupby("month").agg([
    pl.col("funds").sum().alias("total_funds"),
    pl.col("pe").mean().alias("avg_pe")
])

line_config = {
    "columns": ["month", "total_funds", "avg_pe"],
    "row_pivots": ["month"],
    "plugin": "y_line" # Y Line视图
}
```

## 时间序列折线图（Y Line）

```
el LaTeX Tabs Settings Help
Untitled1.ipynb x untitled.py x untitled1.py x Untitled2.ipynb x
Code
pl.col("funds").sum().alias("total_funds"),
pl.col("pe").mean().alias("avg_pe")
])

line_config = {
    "columns": ["month", "total_funds", "avg_pe"],
    "row_pivots": ["month"],
    "plugin": "y_line" # Y Line视图
}

PerspectiveWidget(d1_time.to_pandas(), **line_config)

[ ]: scatter_config = {
    "columns": ["funds", "pe", "ballot", "industry"],
    "row_pivots": [], # 不分组, 以个股为单位
    "aggregates": {"funds": "avg", "pe": "avg", "ballot": "avg"},
    "plugin": "x/y_scatter",
    "x": "funds",
    "y": "pe",
    "color": "industry" # 按行业着色
}

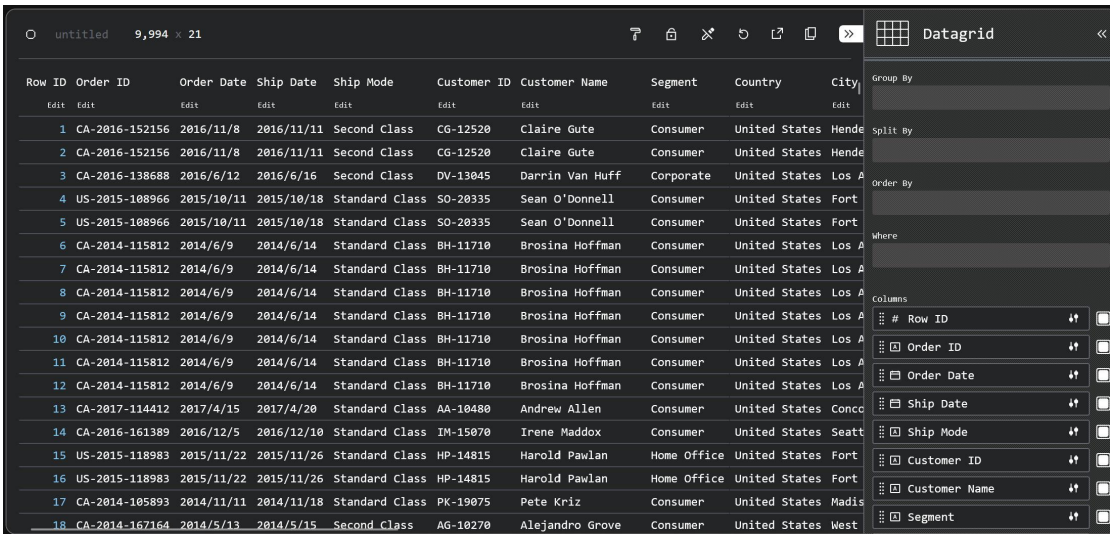
PerspectiveWidget(d1.to_pandas(), **scatter_config)
```

## 操作步骤总结

1. 环境准备：修改 `environment.yml` 添加依赖，运行 `conda env update`。
2. 数据处理：读取 `Parquet` 文件，转换日期类型。
3. 查看数据结构，确认列名和类型。
4. 交互式可视化：初始化 `PerspectiveWidget`，在浏览器中操作界面
5. 设置分组、筛选、排序等选项，导出配置 `JSON`。
6. 自动化配置：将 `JSON` 配置集成到代码中，直接生成所需视图。
7. 探索不同图表类型：使用代码示例创建直方图、折线图、散点图等。

## 注意事项：

数据可视化与交互是非常庞大的话题，在本周的初级教程里，我们专注于低代码探索 (low-code exploration)，所以只介绍了 JupyterLab 和 Perspective。要制作更专业、更精美的可视化解决方案，必然要进一步学习更多的软件框架，甚至更多的计算机语言，以及其他语言的更多的软件框架。总之，还是免不了需要编程的。图形用户界面的其他许多商业软件我们都不推荐，因为不够开放，迟早会限制我们的创新。



The screenshot displays the Perspective Data Grid interface. The main area shows a table with 18 rows and 11 columns. The columns are: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, City, and State. The right sidebar contains a 'Datagrid' panel with various configuration options. The 'Group By' section is empty. The 'Split By' section is empty. The 'Order By' section is empty. The 'Where' section is empty. The 'Columns' section lists the columns and their sort order, with 'Row ID' selected as the first column.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State
1	CA-2016-152156	2016/11/8	2016/11/11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	NV
2	CA-2016-152156	2016/11/8	2016/11/11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	NV
3	CA-2016-138688	2016/6/12	2016/6/16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	CA
4	US-2015-108966	2015/10/11	2015/10/18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Worth	TX
5	US-2015-108966	2015/10/11	2015/10/18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Worth	TX
6	CA-2014-115812	2014/6/9	2014/6/14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	CA
7	CA-2014-115812	2014/6/9	2014/6/14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	CA
8	CA-2014-115812	2014/6/9	2014/6/14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	CA
9	CA-2014-115812	2014/6/9	2014/6/14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	CA
10	CA-2014-115812	2014/6/9	2014/6/14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	CA
11	CA-2014-115812	2014/6/9	2014/6/14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	CA
12	CA-2014-115812	2014/6/9	2014/6/14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	CA
13	CA-2017-114412	2017/4/15	2017/4/20	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	NC
14	CA-2016-161389	2016/12/5	2016/12/10	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	WA
15	US-2015-118983	2015/11/22	2015/11/26	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	TX
16	US-2015-118983	2015/11/22	2015/11/26	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	TX
17	CA-2014-105893	2014/11/11	2014/11/18	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	WI
18	CA-2014-167164	2014/5/13	2014/5/15	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West	CA

