

## 1. 项目背景

本项目基于 Superstore 2015–2018 年的历史销售数据，通过 Python (Pandas、Matplotlib、Seaborn) 进行探索性数据分析 (EDA)，从品类、客户、区域、时间趋势和产品维度识别业务增长点，并为运营策略提供数据支持。

## 2. 数据说明

- 数据条数：9,800
  - 字段数量：18 列
  - 时间范围：2015–2018
  - 主要字段：
    - Order Date
    - Ship Date
    - Customer Segment
    - Region
    - Category / Sub-Category
    - Product Name
    - Sales (GMV 指标)
- 

## 3. 数据清洗与预处理

数据清洗步骤包括：

### (1) 缺失值处理

- Postal Code 存在 11 条缺失值，均来自同一区域，属非关键字段，不影响主分析。
- 未发现严重缺失的问题。

## (2) 格式清洗

- 将 Order Date、Ship Date 转换为 datetime 格式
- 新增 YearMonth 字段用于时间序列分析
- Postal Code 转换为字符串（避免浮点数格式错误）

## (3) 一致性检查

- Segment、Region、Category、Sub-category 均无多种拼写问题
- 数据无明显重复订单

## (4) 异常值处理

无异常值

**总体数据质量良好，可直接用于业务分析。**

---

## 4. 多维度核心分析结果

### 4.1 Category（品类）分析

**销售额结构：**

- **Technology: GMV 最高（核心品类）**
- **Furniture: 第二**
- **Office Supplies: 最低**

**洞察：**

Technology 类高客单价设备推动 GMV，大项目型采购贡献大。

---

### 4.2 Sub-Category（子类别）分析

### ***Top 5 子类别 (高贡献)***

1. Phones
2. Chairs
3. Storage
4. Tables
5. Binders

### ***Bottom 5 子类别 (低贡献)***

1. Fasteners
2. Labels
3. Envelopes
4. Art
5. Supplies

#### **洞察:**

GMV 高度集中在 Phones & 家具类产品；低贡献子类可减少库存压力。

---

### **4.3 Segment (客户类型) 分析**

- **Consumer 占比最高：50.8%** (主力客户)
- Corporate 次之
- Home Office 占比较低

#### **洞察:**

Consumer 是主要 GMV 来源。

---

### **4.4 Region (区域) 分析**

柱状图结果：

- **West: 约 710k → 第一名**
- **East: 约 670k → 第二名**
- **Central: 约 490k → 中等**
- **South: 约 390k → 最弱区域**

**洞察：**

销售资源应重点集中在 West/East, South 为潜力市场。

---

#### 4.5 月度 GMV 趋势（时间序列分析）

折线图显示：

**整体趋势：持续上升（2015 → 2018）**

- 2018 多个月份突破 80k–100k
- 2018 年 12 月达到历史最高（约 120k）

**明显季节性规律：**

- **Q4（10–12 月）为销售旺季**
- **Q1（1–2 月）为低谷**

**洞察：**

典型零售业务周期规律，可提前调整备货与营销节奏。

---

#### 4.6 Top 10 产品分析

前十产品（按 GMV）：

1. **Canon imageCLASS 2200 Advanced Copier (61,599) – 绝对冠军**
2. Fellowes PB500 Electric Punch
3. Cisco TelePresence EX90
4. HON 5400 Series Task Chairs
5. GBC DocuBind TL300
6. GBC Ibimaster 500
7. HP LaserJet 3310
8. HP Designjet T520
9. GBC DocuBind P400
10. Automatic Electric Letter Opener

#### 洞察:

- GMV 高度集中在 **高客单价办公设备**
  - 多数来自 Technology 和 Office Supplies
  - 高价设备是业务的核心驱动力
  - Top10 产品占整体 GMV 非常高
- 

## 5. 综合业务总结

基于 Category、Segment、Region、Time、Product 五个维度，整体结论如下：

1. **业务规模持续增长，呈长期上升趋势**
  2. **GMV 高度依赖高客单价办公设备**（特别是 Technology）
  3. **Consumer 是主力客户，但企业客户更具价值**
  4. **West/East 是绝对核心市场**
  5. **销售明显呈季节性波动（Q4 旺季）**
  6. **产品结构集中度高，重点设备需优先保障供应链**
- 

## 6. 建议 (Actionable Recommendations)

### **(1) 品类方向**

- 增强 Phones、Chairs、Storage 等高贡献品类的库存保障
- 降低低贡献品类（Fasteners 等）库存成本

### **(2) 客户运营**

- Consumer: 适合做促销和新品营销
- Corporate: 推出企业采购套餐，提高稳定收入
- Home Office: 做高客单价产品的精准推荐

### **(3) 区域运营**

- West/East: 持续投入广告、供应链保障
- Central/South: 通过区域折扣、渠道拓展提升渗透率

### **(4) 季节策略**

- Q3 提前备货
- Q4 加大推广力度（旺季）
- Q1 推低价与复购产品平滑淡季

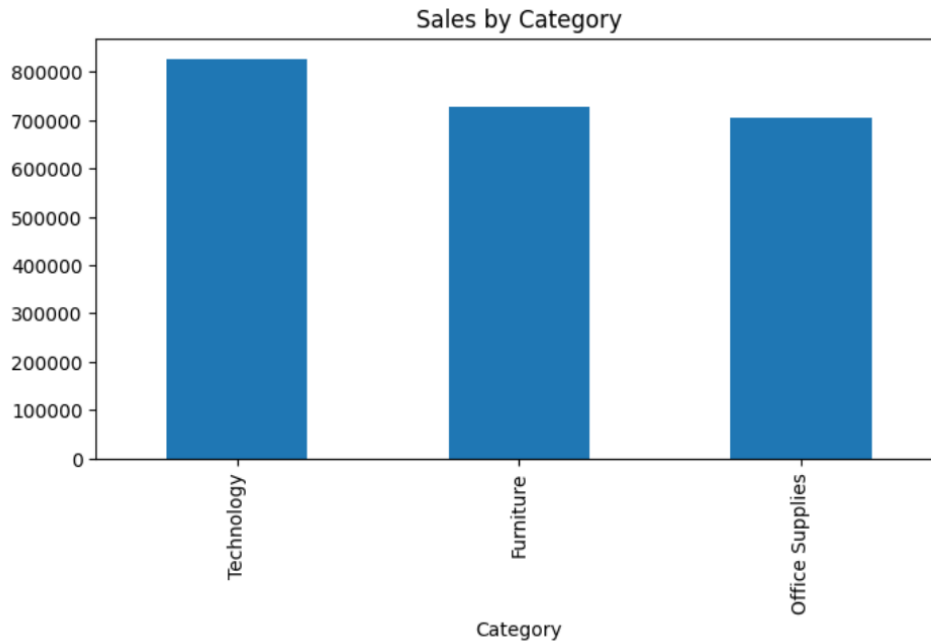
### **(5) 产品策略**

- 优先保障 Top10 高客单价设备的供应链
- 推出延保套餐提高复购与服务收入

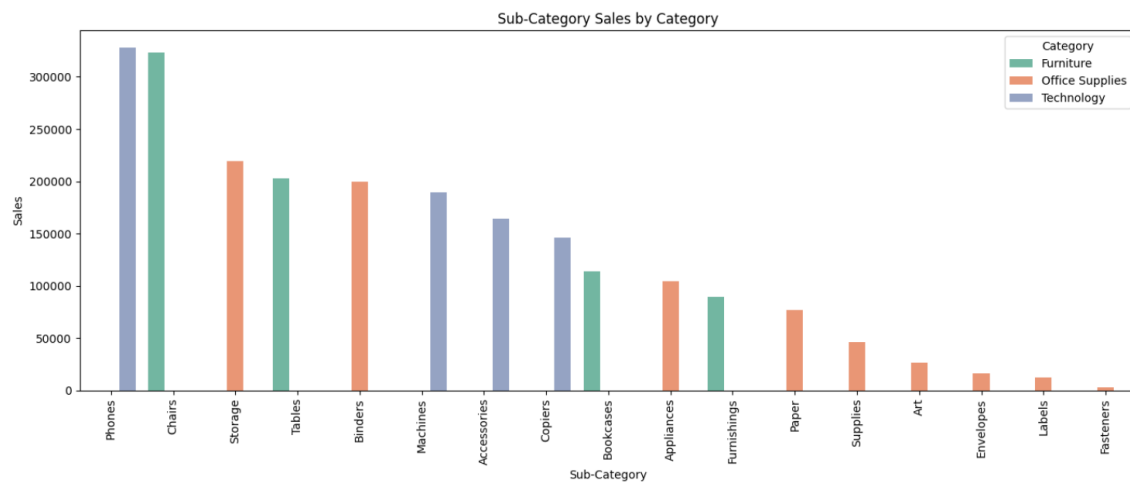
---

## **7. 附录：图表**

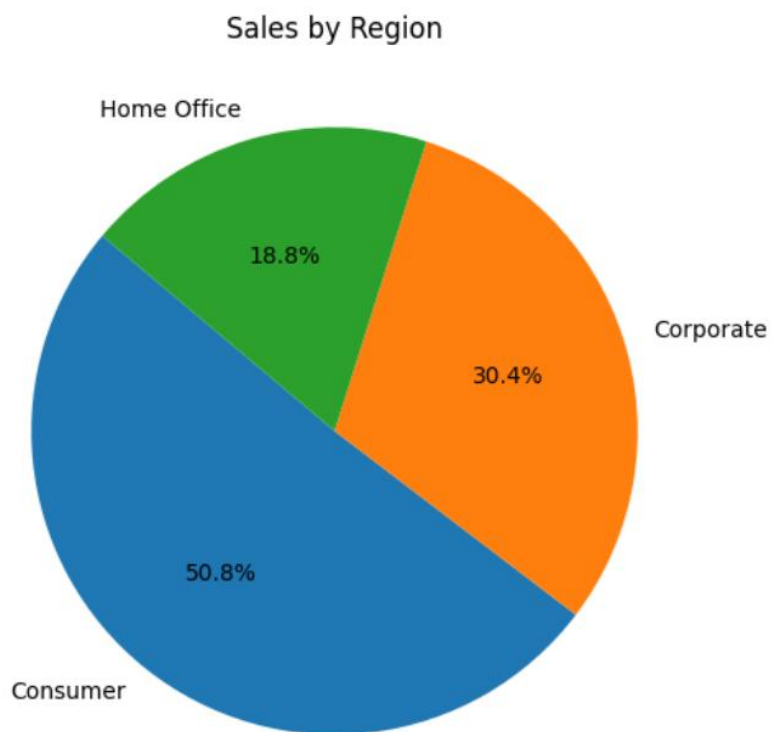
- Category 柱状图



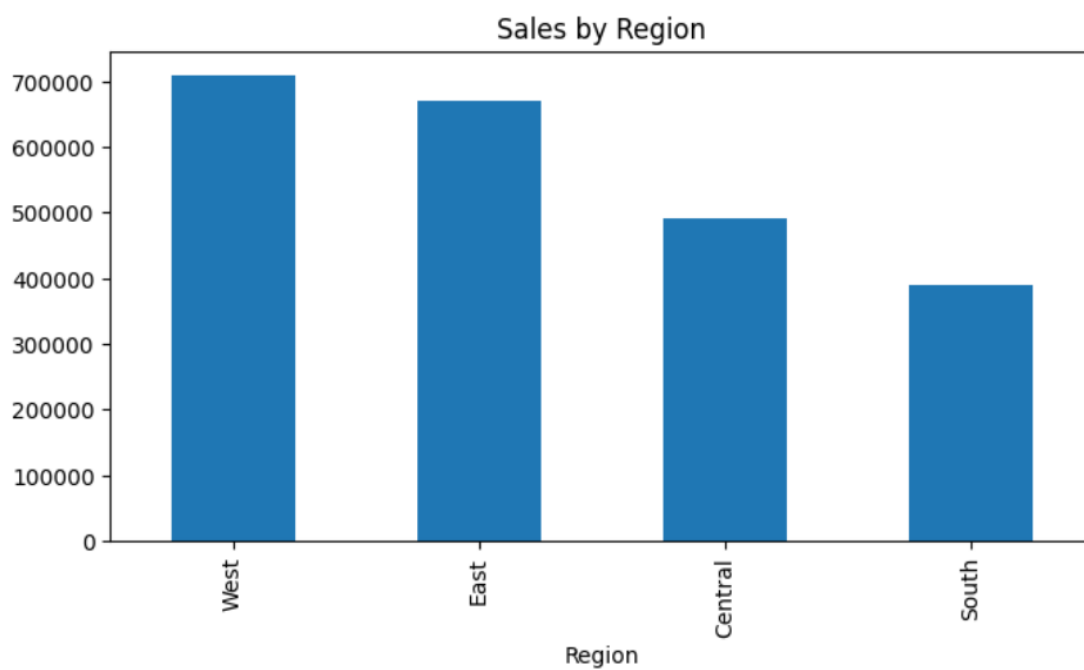
- Subcategory 柱状图



- Segment 饼图

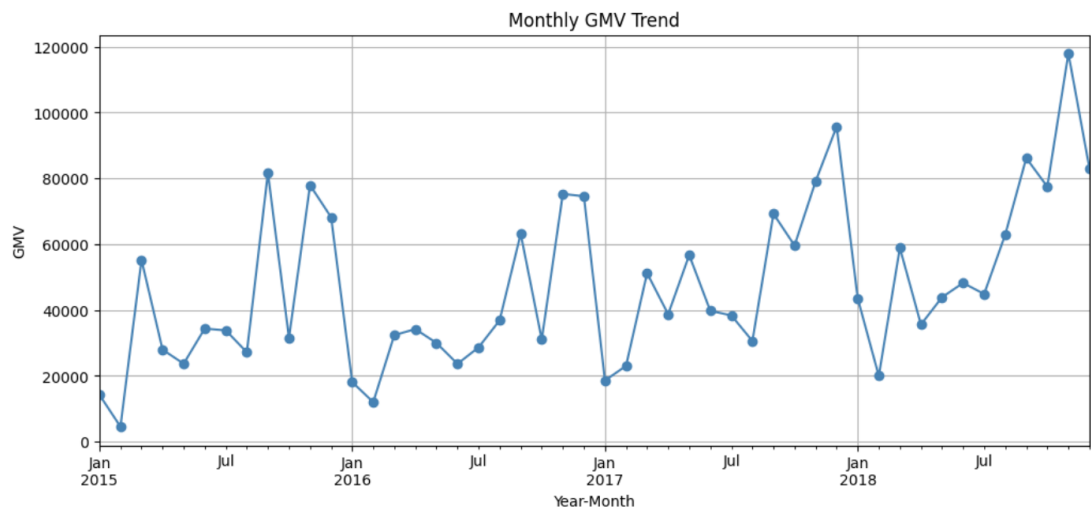


- Region 柱状图





- GMV 月度趋势折线图



- Top10 产品横向柱状图

