

Portfolio Project: Sales Data Cleaning & Reporting

Project Objective:

Clean a messy sales dataset and generate summary reports including:

- Total revenue calculation
- Average order value
- Top 5 selling products
- Monthly revenue trend visualization

Tools Used:

- Python
- Pandas
- Matplotlib

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("sales_data.csv")
df.dropna(inplace=True)
df.drop_duplicates(inplace=True)
df['Order Date'] = pd.to_datetime(df['Order Date'])
df['Total Price'] = df['Quantity Ordered'] * df['Price Each']

total_revenue = df['Total Price'].sum()
avg_order = df['Total Price'].mean()
top_products = df.groupby('Product')['Total Price'].sum().sort_values(ascending=False).head(5)
df['Month'] = df['Order Date'].dt.to_period('M')
monthly_revenue = df.groupby('Month')['Total Price'].sum()

plt.figure(figsize=(10,6))
monthly_revenue.plot(kind='bar', color='orchid')
plt.title("Monthly Revenue")
plt.ylabel("USD")
plt.xlabel("Month")
plt.tight_layout()
plt.savefig("monthly_revenue.png")
df.to_csv("cleaned_sales_data.csv", index=False)
```

Specialization

I specialize in Python-based data services, including data cleaning, preprocessing, and automation.

My focus is on transforming raw, messy datasets into clean, structured formats ready for analysis, visualization, or reporting.

Skilled in:

- CSV/Excel data cleanup
- Pandas & NumPy
- Python scripting for automation
- Data formatting, merging & deduplication
- Visualization with Matplotlib/Seaborn

Looking to work with individuals or businesses needing fast, clean, and professional data solutions.