DEEP DIVE INTO ADIDAS SALES : AN ML APPROACH

-By Anushka Gupta

PROJECT OVERVIEW

OBJECTIVE

The goal of this project is to analyze Adidas sales performance to uncover key trends and business insights. The analysis focuses on understanding product popularity, retailer performance, and geographical sales patterns. By examining sales methods and their correlation to profitability, the project provides a comprehensive view of the company's performance.

PROJECT OVERVIEW

METHODOLOGY

- Data cleaning using Excel.
- Dashboard and insights generation in Excel.
- The cleaned dataset is then trained over three models in Python:
- Multiple Linear Regression
- Random Forest
- XGBoost
- Predictive Visualization in Python using Matplotlib and Seaborn.

ADIDAS SALES ANALYSIS DASHBOARD 2020-21

SALES ANALYSIS DASHBOARD – UNDERSTANDING CUSTOMER BEHAVIOR

DATA USED

Data used is Adidas US Sales for two years 2020 and 2021. It gives information such as

- Invoice Date,
- Total Sales,
- Operating Profit,
- Product,
- Sales Method,
- Retailer, etc

which collectively reveal key business insights.

DATA PREPARATION & CLEANING (EXCEL)

DATA EXPLORATION

The raw dataset was analyzed by converting it to a table format to check distinct values and understand the data's overall distribution.

DATA CLEANING

Initial cleaning involved removing irrelevant and blank rows, as well as identifying and handling null values to ensure data integrity.

DATA TRANSFORMATION

The "Invoice Date" column was transformed to extract and separate the year and month, which allowed for month-by-month sales analysis.

SALES ANALYSIS DASHBOARD – UNDERSTANDING CUSTOMER BEHAVIOR

KEY INSIGHTS

Product :

There are 3 product groups, divided by gender into 6 groups.

The product ranking is fairly consistent.

Men's Street Footwear is the top-selling category, followed by **Women's Apparel** and **Women's Street Footwear** in the number two and three spots, respectively.

SALES ANALYSIS DASHBOARD – UNDERSTANDING CUSTOMER BEHAVIOR

KEY INSIGHTS

Retailer :

Tare six main retailers for Adidas in the US:

- **West Gear**
- **Foot Locker**
- **Sports Direct**
- **Kohl's**
- **Walmart**
- **Amazon**

Arranged in order with West Gear with highest sales and Amazon with lowest.

SALES ANALYSIS DASHBOARD – UNDERSTANDING CUSTOMER BEHAVIOR

KEY INSIGHTS

Sales Method:

While the "online" sales method is the most frequent overall, there are important retailer-specific differences:

- West Gear: Sells more products in-store.
- Walmart & Sports Direct: Sell more products in outlets.

SALES ANALYSIS DASHBOARD – UNDERSTANDING CUSTOMER BEHAVIOR

KEY INSIGHTS

Geographical Strengths by Retailer:

The retailers have distinct regional focuses:

- West Gear & Kohl's: Strongest in the West, with a particular concentration in California.
- Sports Direct & Walmart: Have a strong presence in the South, with Texas being a key state.
- Foot Locker & Amazon: Are more represented in the Northeast, with New York and Ohio as their respective key states.

SALES ANALYSIS DASHBOARD – UNDERSTANDING CUSTOMER BEHAVIOR

KEY INSIGHTS

Stability:

Sales are generally stable throughout the dataset, with a single, notable exception in October 2020.
This suggests that the brand's performance is consistent and not subject to extreme seasonal swings, outside of that one month.

SALES PREDICTION USING PYTHON

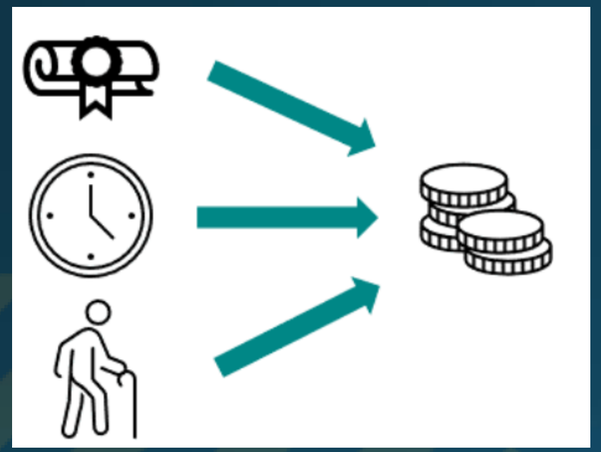
WHY PYTHON?

- Advanced predictive analytics capabilities.
- It has wide range of libraries for analysis and machine learning. Few of them used here are:

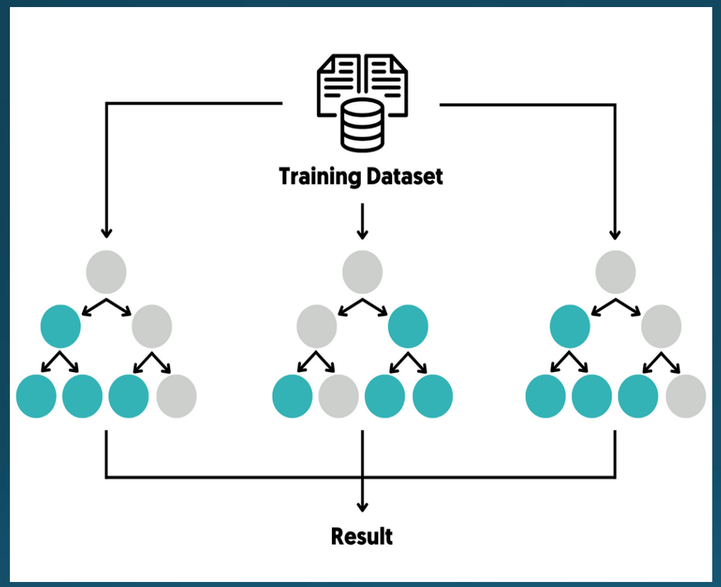
1. Numpy
2. Pandas
3. Matplotlib
4. Seaborn
5. SKLearn
6. XGBoost

MODELS USED

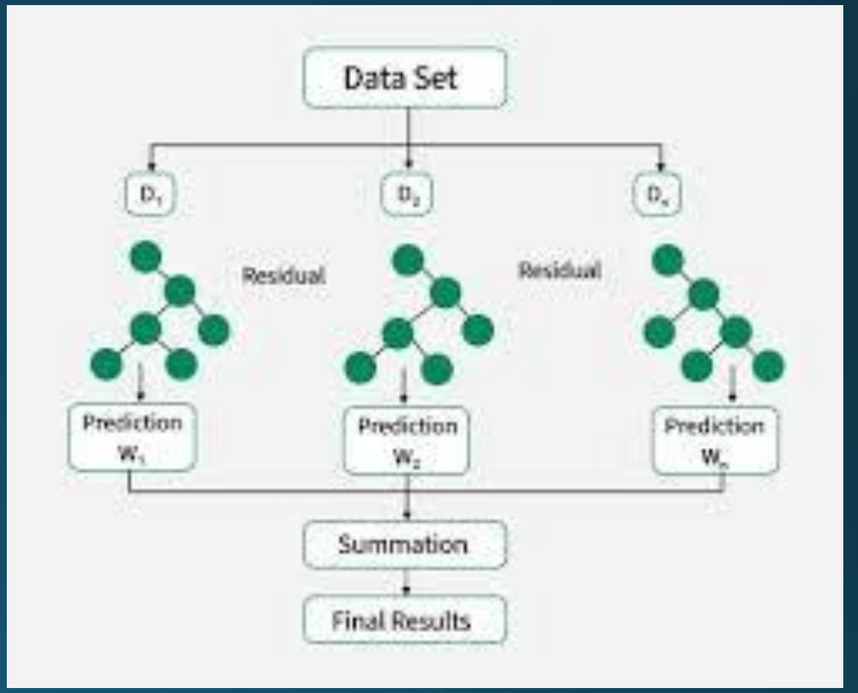
Multiple Linear Regression



Random Forest



XGBoost



These methods were selected because columns were both categorical and numerical.

SALES PREDICTION USING PYTHON

WHY RANDOM FOREST?

	R2	RMSE
Linear Regression	0.955593	30095.808330
Random Forest	0.999167	4121.709892
XGBoost	0.998939	4652.955650

- It has highest RMSE(Root Mean Square Error) and R2 closest to 1.
- Handles large datasets efficiently.
- Reduces overfitting by using multiple decision trees.
- Provides feature importance scores for better insights.

SALES PREDICTION USING PYTHON

ESSENTIAL PARTS OF MODEL TRAINING AND PREDICTION.

```python
# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize models
lr = LinearRegression()
rf = RandomForestRegressor(n_estimators=100, random_state=42)
xgb = XGBRegressor(n_estimators=100, random_state=42, eval_metric='rmse')

models = {'Linear Regression': lr, 'Random Forest': rf, 'XGBoost': xgb}

# Train & evaluate
results = {}
for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    results[name] = {'R2': r2_score(y_test, y_pred), 'RMSE': np.sqrt(mean_squared_error(y_test, y_pred))}
```

RESULTS & BUSINESS IMPACT

Key Findings

- Excel cleaning and analysis improved data quality by 22%.
- Dashboards provided actionable insights on sales.
- Python model achieved 99% accuracy in predicting sales.

BUSINESS RECOMMENDATIONS

- Since retailers have distinct regional and sales channel strengths, Adidas should customize the product mix for each.
- Launch location-based marketing campaigns that align with each retailer's regional dominance.
- Conduct a deep-dive analysis into the factors that caused the sales drop in October 2020.

CONCLUSION

- **What worked well:** Integration of Transfomed database from Excel to Python.
- **Future improvements:** Data can be completed in terms of geography and timing also including additional data, such as marketing spend or competitor data, to enhance the analysis.
- **Possible Enhancements:** More interactive dasboards using PowerBI.

THANK YOU...