

## Лабораторная работа 4. «Аналитика данных в Polymatica Analytics»

### Цель работы:

Освоить основные инструменты платформы Polymatica Analytics для работы с данными, включая создание мультисфер для аналитики. В рамках работы студенты научатся объединять и анализировать данные из различных источников (Excel, CSV, SQL) с помощью Polymatica.

### Оборудование и ПО:

- Установленная платформа Polymatica Analytics <https://poly-a.mgpu.ru/>.
- Данные в файлах Excel и CSV.
- Доступ к базе данных SQL.

### Исходные данные:

1. Excel-файл 1 источник: Финансовые данные компании за 2023 год.
2. Excel-файлы 2 источника: Продажи за 2022 и 2023 годы.
3. CSV-файлы 2 источника: Данные о клиентах и продуктах.
4. Excel и CSV файлы: Финансовые отчёты (Excel) и данные по транзакциям (CSV).
5. SQL-запрос и CSV: Данные из базы данных (SQL-запрос) и CSV-файл с аналитикой.

### Структура лабораторной работы

#### 1. Введение

- Полиматика как инструмент для аналитики данных.
- Описание задач по созданию мультисфер из различных источников данных.
- Важность консолидации данных для качественной аналитики.

#### 2. Цель и задачи

- Цель: научиться создавать мультисферы для аналитики данных из различных источников в Polymatica Analytics.
- Задачи: создание мультисфер из файлов Excel, CSV и базы данных SQL для консолидации и дальнейшего анализа.

### Алгоритм выполнения работы

#### Шаг 1. Загрузка данных и создание мультисферы из одного источника (Excel)

##### 1. Подготовка файла Excel:

- Используйте финансовые данные компании за 2023 год в Excel.
- Проверьте файл на наличие заголовков и корректность данных.

##### 2. Создание мультисферы в Polymatica:

- Войдите в интерфейс Polymatica Analytics.
- Выберите опцию Создать новую мультисферу.
- Загрузите Excel-файл.
- Убедитесь, что заголовки корректно распознаны системой.
- Назначьте типы данных для каждого столбца (например, числовые, текстовые, даты).
- Создайте мультисферу и дайте ей название, например, "Финансовые данные 2023".

##### 3. Проверка мультисферы:

- Проверьте, что данные корректно отображаются в мультисфере.
- Выполните простую агрегацию данных (например, суммирование финансовых показателей).

## **Шаг 2. Создание мультисферы из двух источников (Excel)**

### **1. Подготовка файлов Excel:**

- У вас есть два файла Excel: продажи за 2022 и 2023 годы.
- Убедитесь, что оба файла имеют схожую структуру данных (например, одинаковые заголовки колонок).

### **2. Создание мультисферы:**

- В интерфейсе Polymatica выберите опцию. Создать новую мультисферу.
- Загрузите оба файла Excel.
- Объедините данные по совпадающим полям (например, по продуктам или категориям).
- Проверьте типы данных и выполните агрегацию по продажам за 2022 и 2023 годы.
- Назовите мультисферу "Продажи 2022-2023".

### **3. Проверка данных:**

- Проверьте, что данные из обоих файлов корректно объединены.
- Выполните сравнительный анализ продаж за два года (например, динамику роста).

## **Шаг 3. Создание мультисферы из двух источников (CSV)**

### **1. Подготовка файлов CSV:**

- У вас есть два CSV-файла: данные о клиентах и продуктах.
- Проверьте формат файлов и убедитесь, что данные корректны.

### **2. Создание мультисферы:**

- Выберите Создать мультисферу в интерфейсе Polymatica.
- Загрузите оба CSV-файла.
- Объедините данные по общим полям (например, по идентификатору клиента или продукта).
- Назовите мультисферу "Клиенты и продукты".

### **3. Проверка данных:**

- Проверьте корректность объединения данных из обоих CSV-файлов.
- Выполните агрегацию, например, по количеству покупок на клиента.

## **Шаг 4. Создание мультисферы из Excel и CSV источников**

### **1. Подготовка данных:**

- Используйте один файл Excel с финансовыми данными и один файл CSV с транзакциями.
- Убедитесь, что оба файла имеют пересекающиеся данные (например, идентификаторы транзакций или клиентов).

### **2. Создание мультисферы:**

- Создайте новую мультисферу.
- Загрузите данные из Excel и CSV, объединяя их по общему полю (например, по идентификатору транзакции).
- Проверьте типы данных и выполните агрегацию по нужным полям.

- Назовите мультисферу "Финансы и транзакции".

3. Проверка данных:

- Выполните анализ транзакций в разрезе финансовых показателей.
- Постройте графики для наглядного отображения данных.

**Шаг 5. Создание мультисферы из SQL-запроса и CSV**

1. Подготовка SQL-запроса и CSV файла:

- Напишите SQL-запрос для извлечения данных из базы данных (например, извлечение данных о продажах).
- Подготовьте CSV-файл с дополнительной информацией (например, данные по регионам).

2. Создание мультисферы:

- В интерфейсе Polymatica выберите опцию загрузки данных через SQL-запрос.
- Введите SQL-запрос и загрузите результаты в Polymatica.
- Добавьте CSV-файл и объедините его с данными из SQL-запроса по общему полю.

- Назовите мультисферу "Продажи и регионы".

3. Проверка данных:

- Проверьте, что данные из базы данных и CSV корректно объединились.
- Выполните анализ продаж по регионам и клиентам.

**Отчет.**

- Скриншоты выполнения каждого шага.
- SQL-запрос для выгрузки данных из базы данных.
- Данные Excel и CSV, использованные для создания мультисфер.
- глубокая аналитика с применением статистических методов в виджетах.

## ГЕНЕРАЦИЯ ДАННЫХ

```
```python
import pandas as pd
import sqlite3
```

Создание данных для Excel файла 1: Финансовые данные компании за 2023 год

```
financial_data_2023 = {
    'Month': ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August',
'September', 'October', 'November', 'December'],
    'Revenue': [100000, 120000, 115000, 130000, 125000, 135000, 140000, 150000,
160000, 170000, 165000, 180000],
    'Expenses': [80000, 85000, 90000, 95000, 88000, 92000, 98000, 102000, 105000,
110000, 108000, 115000],
    'Profit': [20000, 35000, 25000, 35000, 37000, 43000, 42000, 48000, 55000, 60000,
57000, 65000]
}
```

Excel файлы 2 источника: Продажи за 2022 и 2023 годы

```
sales_2022 = {
    'Month': ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August',
'September', 'October', 'November', 'December'],
    'Sales': [50000, 60000, 55000, 70000, 75000, 80000, 85000, 90000, 95000, 100000,
105000, 110000]
}
```

```
sales_2023 = {
    'Month': ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August',
'September', 'October', 'November', 'December'],
    'Sales': [55000, 70000, 65000, 80000, 85000, 90000, 95000, 100000, 110000,
115000, 120000, 125000]
}
```

CSV файлы 2 источника: Данные о клиентах и продуктах

```
customers_data = {
    'CustomerID': [101, 102, 103, 104, 105],
    'CustomerName': ['Customer A', 'Customer B', 'Customer C', 'Customer D', 'Customer
E'],
    'Country': ['USA', 'Canada', 'Germany', 'France', 'Japan']
}
```

```
products_data = {
    'ProductID': [201, 202, 203, 204, 205],
    'ProductName': ['Product X', 'Product Y', 'Product Z', 'Product W', 'Product V'],
    'Price': [100, 150, 200, 250, 300]
```

```
}
```

Excel файл для Финансовых отчетов

```
financial_reports = {  
    'ReportID': [1, 2, 3, 4, 5],  
    'ReportType': ['Quarterly', 'Annual', 'Half-Yearly', 'Monthly', 'Annual'],  
    'Year': [2023, 2023, 2023, 2023, 2023],  
    'Revenue': [600000, 1200000, 900000, 300000, 1300000],  
    'Expenses': [400000, 800000, 600000, 200000, 900000],  
    'Profit': [200000, 400000, 300000, 100000, 400000]  
}
```

CSV файл для данных по транзакциям

```
transactions_data = {  
    'TransactionID': [301, 302, 303, 304, 305],  
    'CustomerID': [101, 102, 103, 104, 105],  
    'ProductID': [201, 202, 203, 204, 205],  
    'Quantity': [10, 20, 15, 25, 30],  
    'TotalPrice': [1000, 3000, 3000, 6250, 9000]  
}
```

SQL-запрос и CSV файл с аналитикой

```
sql_data_query = """  
SELECT CustomerID, ProductID, Quantity, TotalPrice  
FROM transactions_data  
WHERE TotalPrice > 5000;  
"""
```

Создание файлов Excel и CSV

```
with pd.ExcelWriter('financial_data_2023.xlsx') as writer:  
    pd.DataFrame(financial_data_2023).to_excel(writer, index=False, sheet_name='2023  
Financial Data')  
    pd.DataFrame(sales_2022).to_excel(writer, index=False, sheet_name='2022 Sales')  
    pd.DataFrame(sales_2023).to_excel(writer, index=False, sheet_name='2023 Sales')  
    pd.DataFrame(financial_reports).to_excel(writer, index=False,  
sheet_name='Financial Reports')
```

```
pd.DataFrame(customers_data).to_csv('customers_data.csv', index=False)  
pd.DataFrame(products_data).to_csv('products_data.csv', index=False)  
pd.DataFrame(transactions_data).to_csv('transactions_data.csv', index=False)
```

Создание базы данных SQLite и выполнение SQL запроса

```
conn = sqlite3.connect(':memory:')  
pd.DataFrame(transactions_data).to_sql('transactions_data', conn, index=False)
```

```

sql_data = pd.read_sql(sql_data_query, conn)
Сохранение SQL данных в CSV файл
sql_data.to_csv('sql_data.csv', index=False)
Вывод данных, чтобы убедиться в корректности
print("Financial Data 2023:\n", pd.DataFrame(financial_data_2023).head())
print("\nSales Data 2022:\n", pd.DataFrame(sales_2022).head())
print("\nSales Data 2023:\n", pd.DataFrame(sales_2023).head())
print("\nCustomer Data:\n", pd.DataFrame(customers_data).head())
print("\nProduct Data:\n", pd.DataFrame(products_data).head())
print("\nTransaction Data:\n", pd.DataFrame(transactions_data).head())
print("\nSQL Data Result:\n", sql_data.head())
```

```

### Варианты заданий

Для всех вариантов провести поиск в сети интернет, в открытых электронных источниках данных, которые соответствуют Вашему варианту. Создать источник данных в виде СУБД(Mysql, Oracle, ClickHouse) или файлы формата CSV, Excel.

**В связи с нестабильной работой облачного сервиса, необходимо провести видеозапись проекта, который разработали и предоставить линк на видеопояснения.**

- **Структурированность:** отчет должен быть четко структурирован с выделенными разделами для каждой части анализа.
- **Четкость визуализации:** все графики и диаграммы должны быть легко интерпретируемыми, с подписями осей и легендами.
- **Детализация фильтров:** подробно опишите, как настраивались фильтры, виджеты, проект и коннектор для анализа данных в Polymatica Dashboards.

#### Вариант 1. Анализ финансовых показателей компании за 2023 год

Используйте данные из Excel-файла с финансовыми показателями компании за 2023 год. Проанализируйте основные статьи доходов и расходов компании. Сформируйте мультисферу с отображением динамики финансовых потоков по месяцам.

#### Вариант 2. Сравнение продаж за 2022 и 2023 годы

Импортируйте данные из Excel-файлов за два года и создайте мультисферу для сравнения объемов продаж. Оцените, как изменились объемы продаж по месяцам и продуктовым категориям.

#### Вариант 3. Сегментация клиентов по регионам

Используйте данные из CSV-файла о клиентах, сегментируйте их по регионам и визуализируйте результаты в Polymatica Analytics. Определите регионы с наибольшей активностью клиентов.

#### Вариант 4. Анализ продаж по продуктовым категориям

На основе данных о продуктах из CSV-файла, выполните анализ продаж по категориям товаров. Постройте диаграммы, отображающие наиболее популярные товарные группы в 2023 году.

### **Вариант 5. Оценка эффективности рекламных кампаний**

Сравните данные по продажам и транзакциям до и после рекламной кампании, используя данные из Excel и CSV-файлов. Выполните глубокую аналитику с применением статистических методов.

### **Вариант 6. Анализ продаж в зависимости от демографических данных клиентов**

Импортируйте данные о клиентах из CSV и выполните анализ продаж в зависимости от возраста и пола. Сформируйте мультисферу, отображающую сегменты клиентов с высокой покупательской активностью.

### **Вариант 7. Анализ данных о транзакциях**

Используйте данные из CSV-файла о транзакциях, чтобы провести анализ по частоте и сумме транзакций. Определите тенденции в покупательской активности клиентов.

### **Вариант 8. Годовой анализ выручки**

Импортируйте данные из Excel-файлов с финансовыми отчетами. Постройте годовой анализ выручки компании за 2022 и 2023 годы. Определите месяцы с наибольшими и наименьшими показателями.

### **Вариант 9. Анализ прибыли по продуктам**

Используйте данные из Excel и CSV о продажах и транзакциях, чтобы оценить прибыльность различных продуктовых категорий. Сформируйте мультисферу для визуализации этого анализа.

### **Вариант 10. Сравнительный анализ клиентов по уровням дохода**

Проанализируйте данные о клиентах, сгруппированных по уровню дохода. Постройте графики, отражающие влияние дохода на покупательские привычки.

### **Вариант 11. Сравнение данных о продажах и транзакциях**

Объедините данные из Excel и CSV-файлов, чтобы сравнить данные о продажах и транзакциях. Найдите закономерности в поведении клиентов, выполнявших крупные покупки.

### **Вариант 12. Анализ региональных продаж**

Используя данные о продажах, выполните анализ по регионам. Определите, какие регионы демонстрируют наибольшую динамику роста продаж.

### **Вариант 13. Оценка динамики продаж новых продуктов**

Импортируйте данные о новых продуктах из CSV-файла и проанализируйте их продажи за 2023 год. Постройте графики, показывающие рост или спад интереса к этим товарам.

### **Вариант 14. Анализ клиентской базы по возрасту**

Сегментируйте клиентскую базу по возрасту и определите возрастные группы, которые приносят наибольшую прибыль компании. Визуализируйте результаты в Polymatica Analytics.

### **Вариант 15. Анализ отказов от покупок**

Используя данные о транзакциях, выполните анализ случаев отмены покупок. Определите, какие факторы могли повлиять на высокие уровни отказов.

### **Вариант 16. Оценка сезонных изменений в продажах**

Импортируйте данные из Excel-файлов и выполните анализ сезонных изменений продаж. Постройте мультисферу, отображающую динамику продаж по кварталам.

### **Вариант 17. Анализ финансовых рисков**

На основе финансовых данных компании проведите анализ рисков. Используйте статистические методы для оценки финансовой устойчивости компании в 2023 году.

### **Вариант 18. Группировка клиентов по активности**

Используйте данные о клиентах и транзакциях для группировки клиентов по уровню активности. Определите категории клиентов с высокой, средней и низкой активностью.

### **Вариант 19. Анализ долгосрочных контрактов**

Импортируйте данные о долгосрочных контрактах компании и выполните анализ доходов от этих контрактов. Постройте прогнозы на 2024 год.

### **Вариант 20. Оценка эффективности каналов продаж**

Сравните данные о продажах по различным каналам (онлайн, офлайн). Визуализируйте, какие каналы принесли наибольшую прибыль в 2023 году.

### **Вариант 21. Анализ кредитной истории клиентов**

Импортируйте данные о кредитных историях клиентов и проанализируйте их связь с покупательской активностью. Определите клиентов с высоким уровнем риска.

### **Вариант 22. Оценка эффективности инвестиций в маркетинг**

Проанализируйте данные по продажам и маркетинговым затратам. Сформируйте мультисферу для оценки рентабельности инвестиций в маркетинговые кампании.

### **Вариант 23. Прогнозирование продаж на 2024 год**

Используя исторические данные по продажам за 2022 и 2023 годы, выполните прогнозирование продаж на 2024 год с помощью статистических методов в Polymatica Analytics.

### **Вариант 24. Анализ частоты покупок клиентов**

Проанализируйте данные о частоте покупок клиентов, используя данные из CSV-файлов. Определите, какие клиенты совершают покупки чаще всего.

### **Вариант 25. Анализ данных по возвратам товаров**

Используйте данные о возвратах товаров из базы данных SQL и CSV-файла. Определите, какие категории товаров имеют наибольший процент возвратов и найдите возможные причины.