

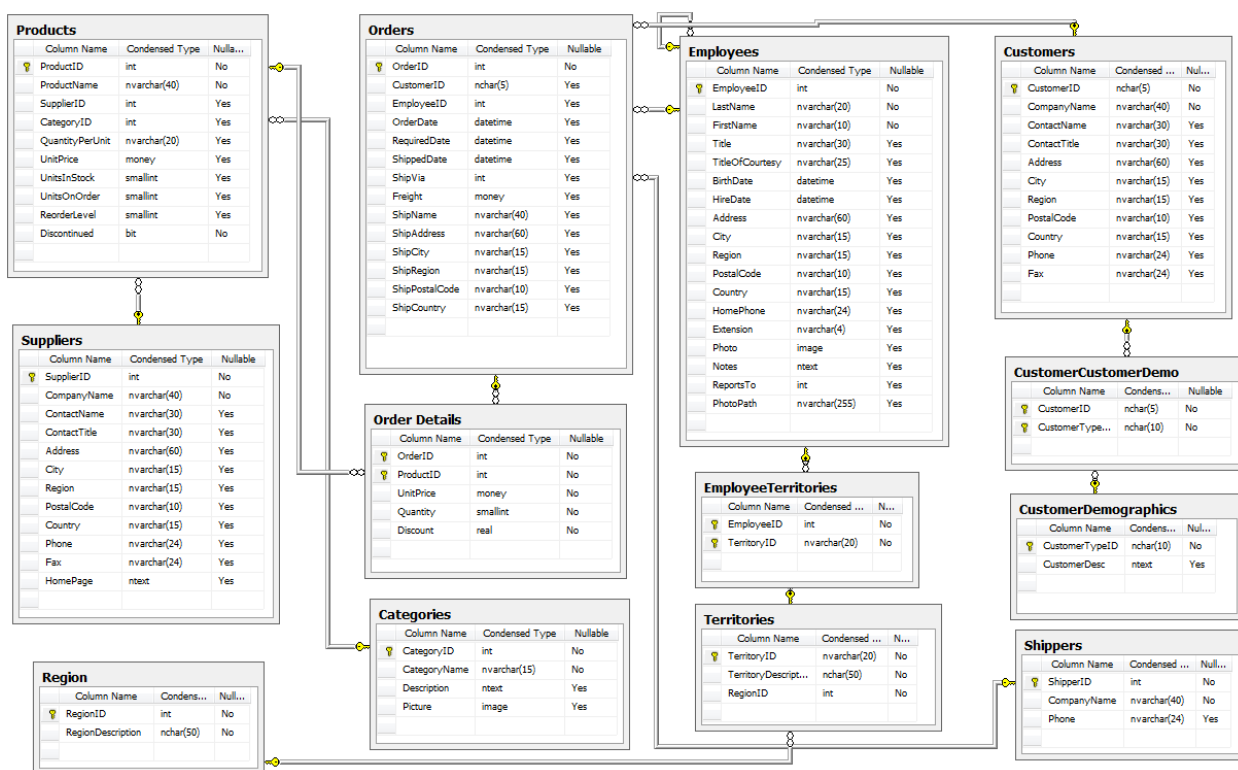
Лабораторная работа № 4

Тема: «Продвинутая визуализация данных на языке Python с помощью библиотеки seaborn»

Цель работы: изучить продвинутые способы работы с библиотекой seaborn.

Теоретическая справка

Диаграммы База данных Northwind sample была предоставлена Microsoft Access в качестве учебной схемы для управления клиентами малого бизнеса, заказами, запасами, закупками, поставщиками, доставкой и сотрудниками. Northwind – иллюстрирует данные ERP-системы малого бизнеса с клиентами, заказами, запасами, закупками, поставщиками, отгрузкой, сотрудниками и единовременным учетом.



SQLite - это библиотека C, предоставляющая облегченную дисковую базу данных, которая не требует отдельного серверного процесса и позволяет получать доступ к базе данных с использованием нестандартного варианта языка запросов SQL. Некоторые приложения могут использовать SQLite для внутреннего хранения данных. Также возможно создать прототип приложения с использованием SQLite, а затем перенести код в более крупную базу данных, такую как PostgreSQL или Oracle.

Для создания подключения к базе в Python необходимо использовать команды:

```
import sqlite3
```

«Визуальная аналитика», Киреев В.С.

```
con = sqlite3.connect("путь к базе на диске")
```

В сочетании с pandas можно обращаться к данным таких баз с помощью команд вида:

```
df=pd.read_sql("SELECT * FROM TABLE_NAME;", con=con)
```

Если в датафрейме содержится столбец с временной меткой (datetime), то данный датафрейм можно проиндексировать по столбцу со временной меткой и изменить гранулярность (периодичность) столбца с количественными данными, командой вида:

```
df.set_index('orderdate')['revenue'].resample('1W').sum()
```

Самостоятельное задание

1. Импортировать данные northwind.db в соответствии с кодом:

```
import sqlite3
import pandas as pd
con=sqlite3.Connection('/content/northwind.db')
df=pd.read_sql("""with revenues as (SELECT
shipregion,
customerid,
orders.orderid,
orderdate,
productid,
(unitprice*quantity*(1-discount)) as revenue
from orders,"order details"
on orders.orderid="order details".orderid
group by orderdate,shipregion,customerid,orders.orderid, productid)
SELECT orderdate,shipregion, customerid,sum(revenue) as revenuedaily,
count(orderid) as ordersdaily
from revenues
group by orderdate;""",con=con)
```

2. С помощью matplotlib и seaborn построить дэшборд по полученным данным, описывающий заказы:
 - 2.1. Создать линейные графики, описывающие историю выручки (revenuedaily) и числа заказов (ordersdaily) (суммарно за неделю)
 - 2.2. Создать круговую диаграмму, описывающую долю региона (shipregion) в сумме выручки (revenue) за весь период
 - 2.3. Создать столбчатую диаграмму, описывающую выручку клиентов, относящихся к классу А. Согласно ABC анализу это список клиентов, которые приносят 80% суммарной выручки за период.
 - 2.4. Создать сетку графиков из 4 вложенных графиков и в каждый из них передать диаграммы из пп.2.1-2.3
3. Изменить на графиках 2.1 тип линии на штрихованную линию.
4. Добавить на график продаж заливку цветом под графиком и изменить цвета (выручка синим цветом, а продажи красным цветом).
5. Добавить заливку фона для графика 2.1 описывающего историю выручки
6. Добавить заголовок для всей сетки, и для каждого графика в отдельности
7. В круговой диаграмме сделать выделение клина, описывающего регион с самой большой долей (выдвинуть его немного, относительно остальных клиньев)
8. Заменить в столбчатой диаграмме палитру базовую на смешанную самостоятельно из зеленого и желтого цвета
9. Изменить размер всей сетки графиков в полтора раза (увеличить и по ширине и по высоте)
10. Добавить сетку с линиями серого цвета на графики 2.1 и 2.3