

داتاکیومنت پروژه پایانی آزمایشگاه پایگاه داده

۴۰۰۷۳۱۲۹ بهزاد فولادی

توضیح دیتابیس

هدف این پروژه، تحلیل شاخص کل بورس ایران در بازه زمانی سال ۲۰۰۸ تا ۲۰۲۳ با استفاده از پایگاه داده و کوئری‌های SQL است. این دیتابیس برای بررسی روند کلی بازار، تغییرات روزانه، نوسانات ماهانه و شناسایی نقاط شکست (Breakout) در شاخص بورس طراحی شده است. پایگاه داده شامل یک جدول اصلی بوده و تمرکز آن بر تحلیل داده‌ها به صورت آماری و زمانی هستش.

کاربردها:

- تحلیل روند بلندمدت شاخص بورس
- محاسبه سود و زیان روزانه
- بررسی نوسان پذیری بازار
- شناسایی نقاط تغییر روند (صعودی و نزولی)

ساختار دیتابیس

ساختار ساده و تحلیلی است و شامل یک موجودیت اصلی می‌باشد:

جدول : Overall_Index

نام ستون	توضیح
DTYYYYMMDD	تاریخ معاملات
OPEN	مقدار شاخص در شروع روز
HIGH	بیشترین مقدار شاخص در روز
LOW	کمترین مقدار شاخص در روز
CLOSE	مقدار شاخص در پایان روز
VOL	حجم معاملات
OPENINT	اطلاعات بهره باز(Open Interest)

از آن جا که هدف پروژه تحلیل داده هاست، رابطه ای بین چند جدول تعریف نشده و تمام تحلیل ها روی همین جدول انجام می شود.

دستورالعمل نصب و راه اندازی پروژه

نرم افزارهای مورد نیاز :

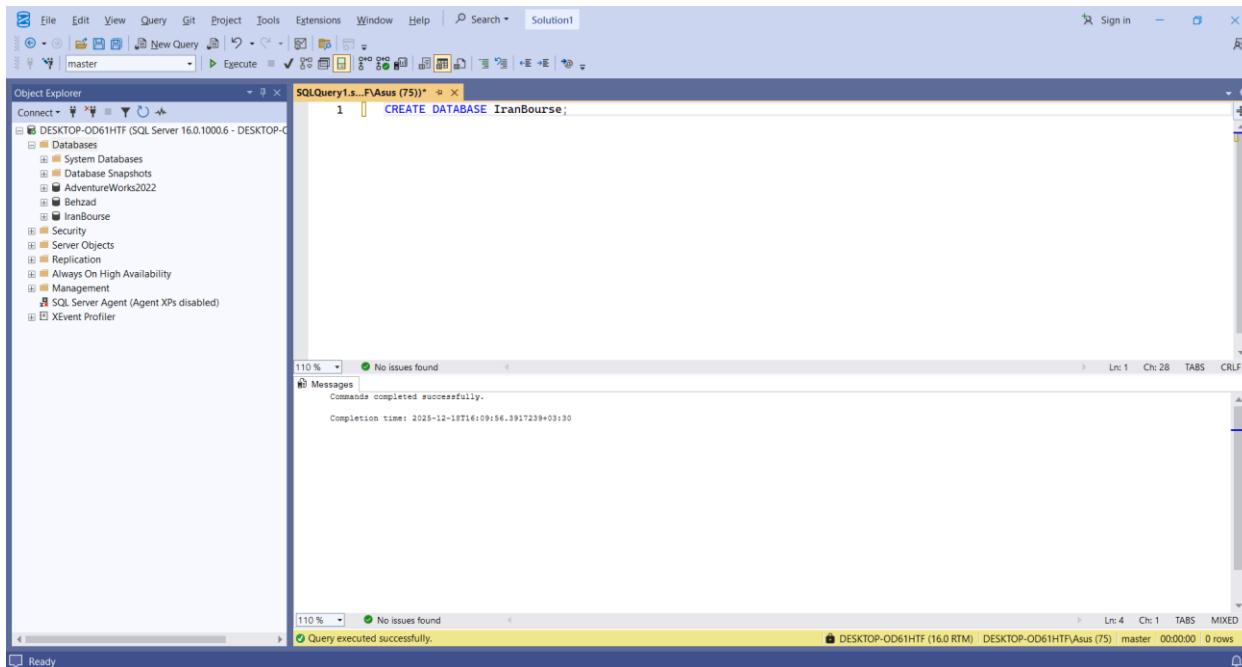
Microsoft SQL Server •

SQL Server Management Studio (SSMS) •

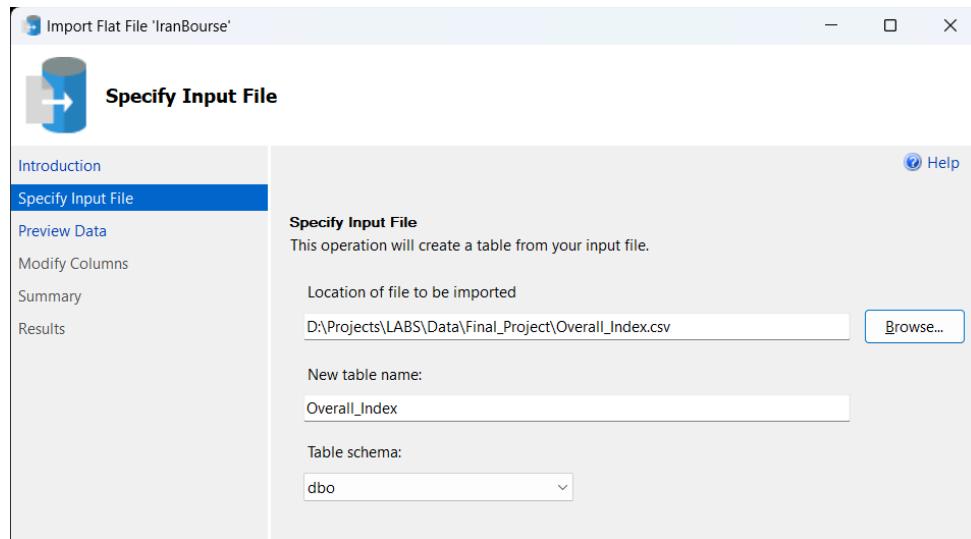
Microsoft Excel •

مراحل راه اندازی

- ۱ . دانلود دیتاست از سایت Kaggle با این لینک (<https://www.kaggle.com/datasets/nimapourmoradi/iran-bourse-overal>) البته در خود فolder docs هم فایل excel موجود هست
- ۲ . باز کردن فایل Excel و تبدیل آن به فرمت CSV
- ۳ . ایجاد یک دیتابیس جدید در SQL Server



۴. استفاده از گزینه Import Flat File (به عکس ها و نکات توجه شود).



The screenshot shows the 'Preview Data' step of the 'Import Flat File' wizard. The sidebar shows the user has moved to the 'Preview Data' step. The main pane displays a preview of the first 50 rows of the CSV file. The columns are labeled DTYYYYMMDD, OPEN, HIGH, LOW, CLOSE, and a timestamp column. The data shows various dates from 20081206 to 20081230 with corresponding price values. Below the preview, there is a note: 'Column names changed due to invalid characters, duplication, etc. Column names can be edited in Modify Columns page.' and a checked checkbox for 'Use Rich Data Type Detection - may provide a closer type fit. However, cells with anomalous values may be dropped.'

DTYYYYMMDD	OPEN	HIGH	LOW	CLOSE	
20081206	9248.8	9248.8	9167.8	9178.3	8
20081207	9178.3	9178.3	9130.3	9130.5	1
20081208	9102.7	9103.4	9079.6	9089.2	1
20081210	9071.6	9071.6	9023.7	9023.7	1
20081213	8973.3	8973	8973	8973.3	3
20081214	8963.9	8964.5	8907	8907	3
20081215	8901.8	8903.6	8890.7	8890.7	1
20081216	8890.8	8890.8	8866.3	8869	1
20081220	8861.5	8861.9	8854.2	8857.1	5
20081221	8857.3	8860.8	8851.6	8851.6	2
20081222	8840.1	8840.6	8832.2	8832.2	1
20081223	8829	8829.5	8803.4	8803.4	1
20081224	8796.4	8802.3	8792.5	8795.7	1
20081227	8784.9	8795.6	8784.9	8786.7	3
20081228	8769.7	8769.7	8757.8	8757.8	2
20081229	8750.1	8751.9	8726.2	8726.9	3
20081230	8717.2	8717.2	8689.8	8695	2
20081231	8691.2	8691.2	8661.2	8661.2	1

The screenshot shows the 'Modify Columns' step of the 'Import Flat File 'IranBourse'' wizard. On the left, a sidebar lists steps: Introduction, Specify Input File, Preview Data, Modify Columns (which is selected), Summary, and Results. The main area is titled 'Modify Columns' and contains a table schema:

Column Name	Data Type	Primary Key	Allow Nulls
DTYYYYMMDD	date	<input type="checkbox"/>	<input type="checkbox"/>
OPEN	float	<input type="checkbox"/>	<input type="checkbox"/>
HIGH	float	<input type="checkbox"/>	<input type="checkbox"/>
LOW	float	<input type="checkbox"/>	<input type="checkbox"/>
CLOSE	float	<input type="checkbox"/>	<input type="checkbox"/>
VOL	NVARCHAR(50)	<input type="checkbox"/>	<input type="checkbox"/>
OPENINT	NVARCHAR(50)	<input type="checkbox"/>	<input type="checkbox"/>
OPENINT2	NVARCHAR(50)	<input type="checkbox"/>	<input type="checkbox"/>
OPENINT3	NVARCHAR(50)	<input type="checkbox"/>	<input type="checkbox"/>

در قسمت Modify Columns باید دیتابایپ ها را مانند تصویر وارد کنید تا در روند اجرای کوئری های به مشکل برخورید (علت استفاده از OPENINT در NVARCHAR ها این است اطلاعات اضافی هستند و کاربرد تحلیلی ندارند)

The screenshot shows the 'Results' step of the 'Import Flat File 'IranBourse'' wizard. On the left, a sidebar lists steps: Introduction, Specify Input File, Preview Data, Modify Columns, Summary, and Results (which is selected). The main area is titled 'Operation Complete' and displays a summary table:

Name	Result
Insert Data	Success

۵. اجرای کوئری‌های SQL ارائه شده برای تحلیل داده‌ها

تمام کوئری‌ها در فolder sql به صورت مستقیم در SSMS قابل اجرا هستند و فقط باید در خود دیتابیس انتخابی Execute شوند.

مثال داده‌ها

: Overall_Index چند رکورد نمونه از جدول

DTYYYYMMDD	OPEN	HIGH	LOW	CLOSE	VOL
20081206	9248.8	9248.8	9167.8	9178.3	8539624
20081207	9180.1	9250.4	9175.6	9223.9	9123456

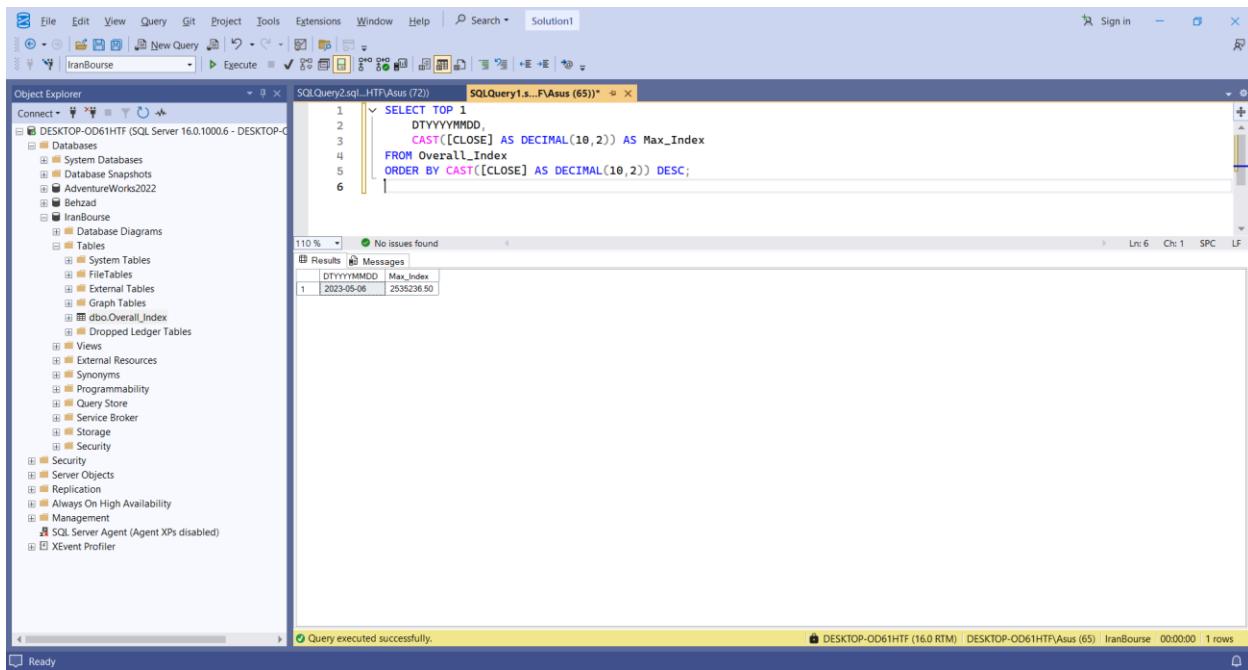
این داده‌ها نمایانگر اطلاعات روزانه شاخص بورس بوده و مبنای تمام تحلیل‌های پروژه هستند.

اجرای کوئری‌های ساده

نمایش کامل رکوردهای جدول شاخص کل:

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. In the Object Explorer, the connection is set to 'DESKTOP-OD61HTF (SQL Server 16.0.1000.6 - DESKTOP-OD61HTF)'. The 'Overall_Index' table is selected in the 'Results' tab of the central pane. The query 'SELECT * FROM Overall_Index;' is displayed above the results. The results grid contains 3587 rows of data, each representing a daily record with columns: DTYYYYMMDD, OPEN, HIGH, LOW, CLOSE, VOL, OPENINT, OPENINT2, and OPENINT3. The data spans from December 2008 to January 2009, showing the daily closing price and volume for the S&P 500 index.

پیدا کردن بالاترین مقدار تاریخی شاخص:



```
SELECT TOP 1
    DTYYYYMMDD,
    CAST([CLOSE] AS DECIMAL(10,2)) AS Max_Index
FROM Overall_Index
ORDER BY CAST([CLOSE] AS DECIMAL(10,2)) DESC;
```

DTYYYYMMDD	Max_Index
2023-05-06	2635236.50

نتیجه‌گیری

در این پژوهش، داده‌های تاریخی شاخص کل بورس ایران از سال ۲۰۰۸ تا ۲۰۲۳ در یک پایگاه داده مبتنی بر SQL Server ذخیره و تحلیل شدند. با استفاده از کوئری‌های ساده و تحلیلی، روند کلی بازار، تغییرات روزانه، نوسانات ماهانه و نقاط مهم تغییر روند شناسایی گردید. نتایج به دست آمده نشان می‌دهد که استفاده از پایگاه داده و زبان SQL ابزار قدرتمندی برای تحلیل داده‌های مالی و تصمیم‌گیری مبتنی بر داده فراهم می‌کند. این پژوهش می‌تواند مبنای مناسب برای تحلیل‌های پیشرفته‌تر و توسعه سیستم‌های تحلیلی در حوزه بازار سرمایه باشد.