

ANALISI DEI DATI



DESCRIZIONE PROGETTO

In un laboratorio chimico è stato introdotto un nuovo macchinario il 1° maggio 2020, che coinvolge le molecole:

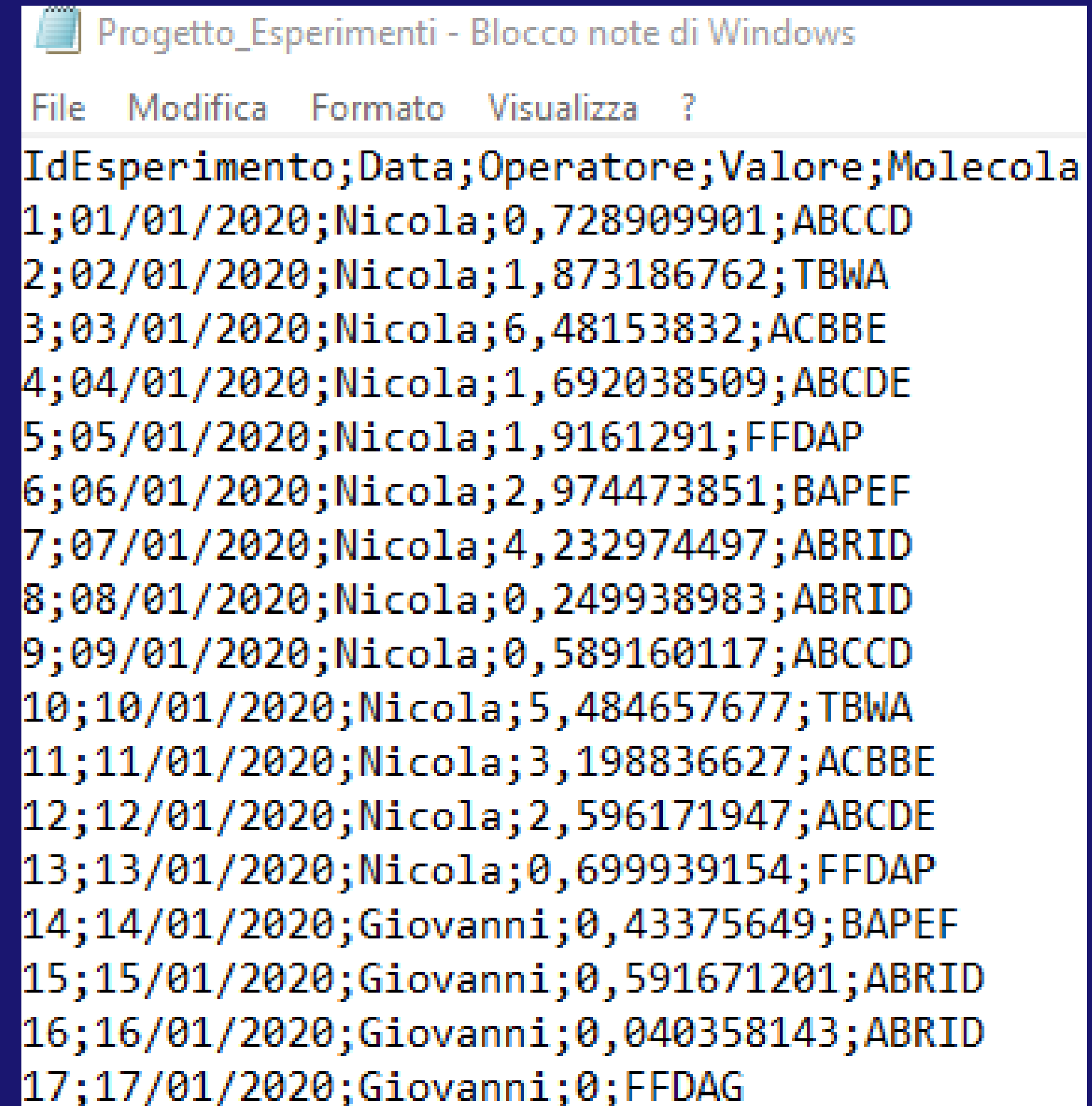
- con nome che inizia per “AB” e finisce con “D”
- con nome che inizia per “F” e non finisce con “P”.

Obiettivo: analizzare come è cambiato il valore medio degli esperimenti per ciascun operatore prima e dopo l'introduzione del macchinario.

DETTAGLI

FILE DI INPUT

- File con estensione csv
- Carattere delimitatore ;
- Presenza di riga con l'intestazione
- Date in formato DD/MM/YYYY
- Numero con carattere , come separatore dei decimali
- 5 colonne, 321 righe + intestazione



```
Progetto_Esperimenti - Blocco note di Windows
File  Modifica  Formato  Visualizza  ?
IdEsperimento;Data;Operatore;Valore;Molecola
1;01/01/2020;Nicola;0,728909901;ABCCD
2;02/01/2020;Nicola;1,873186762;TBWA
3;03/01/2020;Nicola;6,48153832;ACBBE
4;04/01/2020;Nicola;1,692038509;ABCDE
5;05/01/2020;Nicola;1,9161291;FFDAP
6;06/01/2020;Nicola;2,974473851;BAPEF
7;07/01/2020;Nicola;4,232974497;ABRID
8;08/01/2020;Nicola;0,249938983;ABRID
9;09/01/2020;Nicola;0,589160117;ABCCD
10;10/01/2020;Nicola;5,484657677;TBWA
11;11/01/2020;Nicola;3,198836627;ACBBE
12;12/01/2020;Nicola;2,596171947;ABCDE
13;13/01/2020;Nicola;0,699939154;FFDAP
14;14/01/2020;Giovanni;0,43375649;BAPEF
15;15/01/2020;Giovanni;0,591671201;ABRID
16;16/01/2020;Giovanni;0,040358143;ABRID
17;17/01/2020;Giovanni;0;FFDAG
```

PREPARAZIONE

DATI

Dopo l'importazione, ho convertito la colonna Data (formato stringa) in un formato DATE standard, creando la nuova colonna data_converted, necessaria per l'analisi temporale.

Il formato 103 è stato scelto per convertire le date da DD/MM/YYYY, permettendo una valutazione accurata tramite ordinamento, filtro e raggruppamento dei dati.

```
ALTER TABLE dbo.progetto_esperimenti
ADD data_converted DATE;

UPDATE dbo.progetto_esperimenti
SET data_converted =
CONVERT(DATE, Data, 103);
```


STUDIO

COMPARATO

Ho elaborato una query SQL per confrontare, per ogni operatore, la media dei valori degli esperimenti su specifiche molecole, prima e dopo l'introduzione del macchinario.

Utilizzando CTE, ho creato due sottoinsiemi (prima e dopo), ripulito i dati da formati errati (es. 1.234,56), e calcolato le differenze assolute e percentuali tra le medie.

```
with datiprima as
(
select Operatore,AVG(CAST(REPLACE(REPLACE(Valore,'.',''),',','.'))
AS DECIMAL(18,10))) as mediaprima
from dbo.progetto_esperimenti
where data_converted<'2020-05-01' AND ((molecola LIKE 'AB%'
AND molecola LIKE '%D')
or
(molecola LIKE 'F%'
AND molecola NOT LIKE '%P'
))
group by Operatore),
datidopo as
(
select Operatore,AVG(CAST(REPLACE(REPLACE(Valore,'.',''),',','.'))
AS DECIMAL(18,10))) as mediadopo
from dbo.progetto_esperimenti
where data_converted>='2020-05-01' and ((molecola LIKE 'AB%'
AND molecola LIKE '%D')
or
(molecola LIKE 'F%'
AND molecola NOT LIKE '%P'
))
group by Operatore)

SELECT
dp.Operatore,
dp.mediaprima,
dd.mediadopo,
(dd.mediadopo - dp.mediaprima) AS differenzaassoluta,
((dd.mediadopo - dp.mediaprima)/dp.mediaprima) as scostamento
FROM datiprima dp
INNER JOIN datidopo dd ON dp.operatore = dd.Operatore;
```

ANALISI DEI RISULTATI

L'output mostra che per tutti e tre gli operatori si è registrato un incremento nel valore degli esperimenti dal 1 maggio 2020 in poi.

Per Alberto e Nicola l'incremento è stato del 41.9% e 46.5% , mentre per Giovanni del 10.6%

	Operatore	mediaprima	mediadopo	differenzaassoluta	scostamento
1	Alberto	2.0342603905	2.8868315072	0.8525711167	0.419106
2	Giovanni	2.7079465561	2.9942716615	0.2863251054	0.105735
3	Nicola	2.1117036017	3.0945281702	0.9828245685	0.465417