Data Analyst Task

Step 1. Kreiraj tabelu u bazi podataka (u ovom slucaju DB2) i ubaci podatke iz fajla u csv format u tabelu.

DDL tabele:

```sql
CREATE TABLE "FISCS101"."LND_ICE_TEST" (
    "SALESORDERNUMBER" VARCHAR(50 OCTETS),
    "ORDERDATE" DATE,
    "DUEDATE" VARCHAR(50 OCTETS),
    "SHIPDATE" VARCHAR(50 OCTETS),
    "SALES_PERSON" VARCHAR(50 OCTETS),
    "SALES_REGION" VARCHAR(50 OCTETS),
    "SALES_PROVINCE" VARCHAR(50 OCTETS),
    "SALES_CITY" VARCHAR(50 OCTETS),
    "SALES_POSTAL_CODE" VARCHAR(50 OCTETS),
    "CUSTOMER_CODE" VARCHAR(50 OCTETS),
    "CUSTOMER_NAME" VARCHAR(50 OCTETS),
    "CUSTOMER_REGION" VARCHAR(50 OCTETS),
    "CUSTOMER_PROVINCE" VARCHAR(50 OCTETS),
    "CUSTOMER_CITY" VARCHAR(50 OCTETS), |
    "CUSTOMER_POSTAL_CODE" VARCHAR(50 OCTETS),
    "LINEITEM_ID" INTEGER,
    "PRODUCT_CATEGORY" VARCHAR(50 OCTETS),
    "PRODUCT_SUB_CATEGORY" VARCHAR(50 OCTETS),
    "PRODUCT_NAME" VARCHAR(50 OCTETS),
    "PRODUCT_CODE" VARCHAR(50 OCTETS),
    "UNIT_COST" DECIMAL(13 , 4),
    "UNITPRICE" DECIMAL(13 , 4),
    "UNITPRICEDISCOUNT" DECIMAL(13 , 4),
    "ORDERQTY" INTEGER,
    "UNIT_FREIGHT_COST" DECIMAL(13 , 4)
)
```

Create_tab.sql

Step 2. Upitom u tabelu moze se videti struktura podataka.

	SALESORDERNUMBER	ORDERDATE	DUEDATE	SHIPDATE	SALES_PERSON	SALES_REGION	SALES_PROVINCE	SALES_CITY	SALES_POSTAL_CODE	CUSTOMER_CODE	CUSTOMER_NAME	CUSTOMER_REGION	CUSTOM
1	SO71774	2008-06-01 00:00.0		00:00.0	Linda Mitchell	United States	Utah	Nevada	84407	AW00029847	Good Toys	United States	California
2	SO71774	2008-06-01 00:00.0		00:00.0	Linda Mitchell	United States	Utah	Nevada	84407	AW00029847	Good Toys	United States	California
3	SO71775	2008-06-01 00:00.0		00:00.0	Tete Mensa-Annan	United States	Massachusetts	Cambridge	2139	AW00030031	Cycle Clearance	United States	Oregon
4	SO71775	2008-06-01 00:00.0		00:00.0	Tete Mensa-Annan	United States	Massachusetts	Cambridge	2139	AW00030031	Cycle Clearance	United States	Oregon
5	SO71775	2008-06-01 00:00.0		00:00.0	Tete Mensa-Annan	United States	Massachusetts	Cambridge	2139	AW00030031	Cycle Clearance	United States	Oregon
6	SO71777	2008-06-01 00:00.0		00:00.0	David Campbell	United States	Washington	Bellevue	98004	AW00030064	Demand Distributors	United States	Washingto
7	SO71777	2008-06-01 00:00.0		00:00.0	David Campbell	United States	Washington	Bellevue	98004	AW00030064	Demand Distributors	United States	Washingto
8	SO71778	2008-06-01 00:00.0		00:00.0	Jillian Carson	United States	Minnesota	Duluth	55802	AW00029501	Purchase Mart	United States	Massachus
9	SO71778	2008-06-01 00:00.0		00:00.0	Jillian Carson	United States	Minnesota	Duluth	55802	AW00029501	Purchase Mart	United States	Massachus

CUSTOMER_POSTAL_CODE	LINEITEM_ID	PRODUCT_CATEGORY	PRODUCT_SUB_CATEGORY	PRODUCT_NAME	PRODUCT_CODE	UNIT_COST	UNITPRICE	UNITPRICEDISCOUNT	ORDERQTY	UNIT_FREIGHT_COST
95603	110,562	Components	Road Frames	ML Road Frame-W - Yellow, - FR-R72Y-48	451.1785	594.83	0	1	11.0043
95603	110,563	Components	Road Frames	ML Road Frame-W - Yellow, : FR-R72Y-38	451.1785	594.83	0	1	11.0043
97321	110,564	Bikes	Touring Bikes	Touring-1000 Blue, 46	BK-T79U-46	1,852.4223	2,384.07	0	1	17.4444
97321	110,565	Components	Brakes	Front Brakes	FB-9873	59.1075	106.5	0	3	17.4444
97321	110,566	Clothing	Jerseys	Short-Sleeve Classic Jersey, L	SJ-0194-L	51.9653	53.99	0	4	17.4444
98801	110,568	Components	Handlebars	HL Road Handlebars	HB-R956	66.7498	120.27	0	3	14.3418
98801	110,569	Components	Road Frames	ML Road Frame-W - Yellow, - FR-R72Y-48	451.1785	594.83	0	2	14.3418
2093	110,570	Clothing	Gloves	Half-Finger Gloves, L	GL-H102-L	11.4491	24.49	0	1	8.5066
2093	110,571	Bikes	Touring Bikes	Touring-3000 Blue, 58	BK-T18U-58	576.806	742.35	0	1	8.5066
2093	110,572	Components	Handlebars	HL Touring Handlebars	HB-T928	50.8213	91.57	0	1	8.5066
2093	110,573	Components	Touring Frames	HL Touring Frame - Blue, 46	FR-T98U-46	752.1796	1,003.91	0	1	8.5066
80030	110,574	Bikes	Mountain Bikes	Mountain-400-W Silver, 42	BK-M38S-42	524.723	769.49	0	2	26.0857

Step 3. Posto je zahtev od korisnika da se analizira zasto je doslo do smanjenja profita poslednjih nekoliko meseci, moze se napraviti nekoliko upita.

```sql
SELECT
    SALES_PERSON,
    count(1)
FROM
    FISCS101.LND_ICE_TEST
GROUP BY SALES_PERSON
```

SALES_PERSON	2
Amy Alberts	97
David Campbell	2,247
Jillian Carson	7,825
Linda Mitchell	7,107
Michael Blythe	7,069
Pamela Ansman-Wolfe	2,064
Rachel Valdez	1
Ranjit Varkey Chudukatil	2,977
Shu Ito	4,545
Stephen Jiang	642
Tete Mensa-Annan	1,893
Tsvi Reiter	5,417

```
SELECT
    CUSTOMER_NAME,
    count(1)
FROM
    FISCS101.LND_ICE_TEST
GROUP BY CUSTOMER_NAME;
```

	CUSTOMER_NAME	2
1	A Bike Store	73
2	A Great Bicycle Company	14
3	A Typical Bike Shop	75
4	Accessories Network	4
5	Acclaimed Bicycle Company	11
6	Active Cycling	2
7	Active Systems	6
8	Active Transport Inc.	71
9	Activity Center	73
10	Advanced Bike Components	394
11	Aerobic Exercise Company	10
12	Affordable Sports Equipment	269
13	All Cycle Shop	28
14	All Seasons Sports Supply	12
15	Alpine Ski House	14
16	Alternative Vehicles	11
17	Another Bicycle Company	176
18	Area Bike Accessories	273
19	Associated Bikes	16
20	Atypical Bike Company	2
21	Authorized Bike Sales and Rental	82
22	Basic Bike Company	27
23	Basic Sports Equipment	64
24	Beneficial Exercises and Activities	5
25	Best o' Bikes	34
26	Better Bike Shop	359
27	Bicycle Accessories and Kits	11
28	Bicycle Lines Distributors	250

… do 420 rekorda.

Step 4. Napraviti nekoliko sumarnih izvestaja po datumu I drugim kriterijumima, u ovom slucaju kako se kretala prodaja za svakog prodavca (sales_person).

```
SELECT
IFNULL(YEAR,'GRAND TOTAL') AS YEAR,
IFNULL(MONTH,'TOTAL FOR YEAR') AS MONTH,
IFNULL(SALES_PERSON,'Total FOR SALES_PERSON') AS SALES_PERSON,
total,
sum_amt
FROM (
SELECT
    YEAR (orderdate) AS YEAR,
    MONTH  (orderdate) AS MONTH,
    SALES_PERSON,
    sum(UNITPRICE*ORDERQTY) total,
    TO_CHAR(NVL(SUM(UNITPRICE*ORDERQTY), 0), '999,999,999,999.99') AS sum_amt
    FROM
    FISCS101.LND_ICE_TEST
    GROUP BY ROLLUP (
    YEAR (orderdate),
    MONTH  (orderdate), SALES_PERSON)
ORDER BY
SALES_PERSON,
sum(UNITPRICE*ORDERQTY) DESC,
YEAR,
MONTH
)
;
```

	YEAR	MONTH	SALES_PERSON	TOTAL	SUM_AMT
1	2,007	2	Amy Alberts	44,008.23	44,008.23
2	2,007	11	Amy Alberts	35,992.28	35,992.28
3	2,007	8	Amy Alberts	32,750.98	32,750.98
4	2,007	12	Amy Alberts	23,231.45	23,231.45
5	2,006	11	Amy Alberts	16,266.89	16,266.89
6	2,008	2	Amy Alberts	2,574.95	2,574.95
7	2,007	3	Amy Alberts	2,443.35	2,443.35
8	2,007	9	Amy Alberts	2,443.35	2,443.35
9	2,006	7	David Campbell	364,605.92	364,605.92
10	2,007	9	David Campbell	337,857.3	337,857.30
11	2,007	8	David Campbell	334,531.39	334,531.39
12	2,007	7	David Campbell	289,434.44	289,434.44
13	2,006	10	David Campbell	280,144.9	280,144.90
14	2,007	4	David Campbell	276,115.53	276,115.53
15	2,006	8	David Campbell	259,819.05	259,819.05
16	2,006	11	David Campbell	258,449.62	258,449.62
17	2,007	12	David Campbell	241,005.35	241,005.35
18	2,008	6	David Campbell	237,942.59	237,942.59
19	2,008	4	David Campbell	217,149.29	217,149.29
20	2,006	2	David Campbell	212,681.79	212,681.79
21	2,007	1	David Campbell	200,088.67	200,088.67
22	2,007	10	David Campbell	198,067.85	198,067.85
23	2,006	5	David Campbell	196,617.83	196,617.83
24	2,007	5	David Campbell	189,980.62	189,980.62
25	2,008	5	David Campbell	188,786.54	188,786.54
26	2,005	11	David Campbell	188,222.95	188,222.95
27	2,008	3	David Campbell	184,134.06	184,134.06
28	2,005	10	David Campbell	179,205.91	179,205.91

… do 365 rekorda.

Ako se uzme za kriterijum customer_name, onda je slicno:

```sql
=SELECT
 IFNULL(YEAR,'GRAND TOTAL') AS YEAR,
 IFNULL(MONTH,'TOTAL FOR YEAR') AS MONTH,
 IFNULL(CUSTOMER_NAME,'Total FOR SALES_PERSON') AS CUSTOMER_NAME,
 total,
 sum_amt
 FROM (
 SELECT
     YEAR (orderdate) AS YEAR,
     MONTH  (orderdate) AS MONTH,
     CUSTOMER_NAME,
     sum(UNITPRICE*ORDERQTY) total,
     TO_CHAR(NVL(SUM(UNITPRICE*ORDERQTY), 0), '999,999,999,999.99') AS sum_amt
     FROM
     FISCS101.LND_ICE_TEST
     GROUP BY ROLLUP (
     YEAR (orderdate),
     MONTH  (orderdate), CUSTOMER_NAME)
 ORDER BY
 CUSTOMER_NAME,
 sum(UNITPRICE*ORDERQTY) DESC,
 YEAR,
 MONTH
 )
 ;
```

YEAR	MONTH	CUSTOMER_NAME	TOTAL	SUM_AMT
2,006	2	A Bike Store	48,418.65	48,418.65
2,006	5	A Bike Store	45,077.75	45,077.75
2,005	11	A Bike Store	34,599.82	34,599.82
2,005	8	A Bike Store	19,063.7	19,063.70
2,005	9	A Great Bicycle Company	4,508.93	4,508.93
2,005	12	A Great Bicycle Company	3,806.96	3,806.96
2,006	3	A Great Bicycle Company	3,806.96	3,806.96
2,006	6	A Great Bicycle Company	3,131.96	3,131.96
2,007	3	A Great Bicycle Company	782.99	782.99
2,007	12	A Great Bicycle Company	62.09	62.09
2,006	6	A Typical Bike Shop	50,933.25	50,933.25
2,005	12	A Typical Bike Shop	49,062.88	49,062.88
2,006	3	A Typical Bike Shop	37,334.52	37,334.52
2,005	9	A Typical Bike Shop	7,432.22	7,432.22
2,008	6	Accessories Network	2,394.05	2,394.05
2,007	9	Accessories Network	742.35	742.35
2,007	12	Accessories Network	742.35	742.35
2,007	9	Acclaimed Bicycle Company	7,900.41	7,900.41
2,007	12	Acclaimed Bicycle Company	3,531.66	3,531.66
2,008	3	Acclaimed Bicycle Company	1,957.2	1,957.20
2,008	6	Acclaimed Bicycle Company	1,003.91	1,003.91
2,006	12	Active Cycling	2,348.97	2,348.97
2,007	3	Active Cycling	660.12	660.12
2,007	4	Active Systems	576.88	576.88
2,006	10	Active Systems	449.94	449.94
2,006	7	Active Systems	69.98	69.98
2,006	5	Active Transport Inc.	50,467.83	50,467.83
2,005	11	Active Transport Inc.	44,997.95	44,997.95

… do 2666 rekorda.

Dalje, mogu se napraviti izvestaji po ostalim kriterijumima I analizirati kada opada prodaja u poslednjim mesecima.

Izraz tipa IFNULL(YEAR1, 'GRAND TOTAL') YEAR1 sluzi da bi se dobio total za mesec, godinu I grand total.

YEAR	MONTH	SALES_PERSON	TOTAL	SUM_AMT
2,005	7	Tsvi Reiter	177,144.6	177,144.60
[NULL]	[NULL]	Total	1,122,105.185	101,122,105.1!
2,007	[NULL]	Total	3,835,579.785	38,835,579.79
2,006	[NULL]	Total	33,265,946.06	33,265,946.06
2,008	[NULL]	Total	17,861,371.53	17,861,371.53
2,005	[NULL]	Total	11,159,207.81	11,159,207.81
2,007	8	Total	5,196,945.77	5,196,945.77
2,006	8	Total	4,957,940.98	4,957,940.98
2,007	9	Total	4,436,861.38	4,436,861.38
2,006	11	Total	4,260,108.925	4,260,108.93
2,007	11	Total	4,086,880.09	4,086,880.09
2,007	5	Total	4,074,109.245	4,074,109.25
2,008	5	Total	4,007,558.45	4,007,558.45
2,006	9	Total	3,789,684.71	3,789,684.71
2,008	6	Total	3,455,721.7	3,455,721.70
2,007	2	Total	3,380,022.105	3,380,022.11
2,007	12	Total	3,354,296.61	3,354,296.61

Step 5. Napraviti upit za izvestaj koji ce da prikaze rezultat gde je razlika maksimalnog I minimalnog prihoda veca ili jednaka 90%. Opet se to moze menjati po raznim kriterijumima, u ovom slicaju je sales_person. Mogu se dodatno ubaciti filteri za upit po datumu.

```sql
SELECT
SALES_PERSON,
max(amount) AS max,
min(amount) AS min,
100 - (min(amount)*100/max(amount))  percentage_drop
FROM
(
WITH test AS (
SELECT
    SALESORDERNUMBER,
    ORDERDATE,
    TO_CHAR(orderdate,'YYYY') AS YEAR1,
    TO_CHAR (orderdate, 'MM') AS MONTH1,
    DUEDATE,
    SHIPDATE,
    SALES_PERSON,
    SALES_REGION,
    SALES_PROVINCE,
    SALES_CITY,
    SALES_POSTAL_CODE,
    CUSTOMER_CODE,
    CUSTOMER_NAME,
    CUSTOMER_REGION,
    CUSTOMER_PROVINCE,
    CUSTOMER_CITY,
    CUSTOMER_POSTAL_CODE,
    LINEITEM_ID,
    PRODUCT_CATEGORY,
    PRODUCT_SUB_CATEGORY,
    PRODUCT_NAME,
    PRODUCT_CODE,
    UNIT_COST,
    UNITPRICE,
    UNITPRICEDISCOUNT,
    ORDERQTY,
    UNIT_FREIGHT_COST
FROM
    FISCS101.LND_ICE_TEST)
SELECT
    IFNULL(YEAR1, 'GRAND TOTAL') YEAR1,
    IFNULL(MONTH1, 'Total for year') MONTH1,
    IFNULL(SALES_PERSON, 'Total for month') SALES_PERSON,
    sum(UNITPRICE * ORDERQTY) amount,
    TO_CHAR(NVL(SUM(UNITPRICE * ORDERQTY), 0), '999,999,999,999.99') AS amount_frm
FROM
    test
GROUP BY
    ROLLUP (
    YEAR1 ,
    MONTH1 ,
    SALES_PERSON)
ORDER BY
    SALES_PERSON,
    YEAR1,
    MONTH1,
    sum(UNITPRICE * ORDERQTY) DESC
    )
GROUP BY
SALES_PERSON
having 100 - (min(amount)*100/max(amount)) >= 90
ORDER BY
100 - (min(amount)*100/max(amount)) DESC,
SALES_PERSON
;
```

SALES_PERSON	MAX	MIN	PERCENTAGE_DROP
Linda Mitchell	1,169,783.96	6,522.07	100
Total for month	101,122,105.185	638,871.55	100
David Campbell	364,605.92	5,173.84	99
Stephen Jiang	190,039.95	2,357.08	99
Pamela Ansman-Wolfe	488,240.19	24,245.85	96
Ranjit Varkey Chudukatil	1,108,757.57	50,090.88	96
Tete Mensa-Annan	395,877.255	16,139.935	96
Amy Alberts	44,008.23	2,443.35	95
Shu Ito	751,058.975	59,074.1	93
Jillian Carson	811,772.235	82,173.32	90

Preporuka za dalju analizu: u razgovoru sa biznis grupom, videti koja pitanja oni imaju u vezi prodaje I koje podatke oni preporucuju da se analiziraju. U skladu sa tim napraviti nove upite I prezentovati izvestaje.

SQL nije jezik koji je specijalizovan za dobijanje statistickih izvestaja, mada ima nekoliko ugradjenih statistickih funkcija. U ovom primeru su koriscene analiticke (OLAP) funkcije.

Ovim se zakljucuje 6-8 sati predvidjenih za ovu analizu. Posto nemam instaliran Tableau, to bi zahtevalo vise vremena. Svrha Tableau I ostalih BI alata je da vizuelno prezentuju rezultate upita u obliku tabela, grafika I dashborda. Oni se oslanjaju na unapred uradjene upite (query), kao I na vec pripremljene sumarne podatke, koji su rezultati tih upita. Mogu da koriste razlicite filtre koji kastomizuju upite.