Unit 2-4 Exercises

1. Using Formatted Input

The raw data file **`sales1.dat`** has employee information for the Australian and U.S. sales staff.
The record layout is shown in the table below:

Layout for **`sales1.dat`**

	Field Description	Starting Column	Length of Field	Data Type
➡	Employee ID	1	6	Numeric
	First Name	8	12	Character
➡	Last Name	21	18	Character
	Gender	40	1	Character
➡	Job Title	43	20	Character
➡	Salary	64	8	Numeric $100,000
	Country	73	2	Character 'AU' or 'US'
	Birth Date	76	10	Numeric mm/dd/yyyy
➡	Hire Date	87	10	Numeric mm/dd/yyyy

a. Create a new SAS data set named **`sales_staff`** that contains the fields indicated by arrows in the
layout table.

b. Print **`sales_staff`** and add an appropriate title.

Partial PROC PRINT Output (165 Total Observations)

Australian and US Sales Staff

Employee_ ID	Last_Name	Job_Title	Salary	Hire_ Date
120102	Zhou	Sales Manager	108255	10744
120103	Dawes	Sales Manager	87975	5114
120121	Elvish	Sales Rep. II	26600	5114
120122	Ngan	Sales Rep. II	27475	6756
120123	Hotstone	Sales Rep. I	26190	17440
120124	Daymond	Sales Rep. I	26480	17226

2. Using Formatted Input and the Subsetting IF Statement

The raw data file **`sales1.dat`** has employee information for the Australian and U.S. sales staff. The record layout is shown in the table below:

Layout for **`sales1.dat`**

	Field Description	Starting Column	Length of Field	Data Type
➡	Employee ID	1	6	Numeric
	First Name	8	12	Character
➡	Last Name	21	18	Character
	Gender	40	1	Character
➡	Job Title	43	20	Character
➡	Salary	64	8	Numeric $100,000
➡	Country	73	2	Character 'AU' or 'US'
	Birth Date	76	10	Numeric mm/dd/yyyy
➡	Hire Date	87	10	Numeric mm/dd/yyyy

a. Create two SAS data sets from the raw data file, and base them on the country of the trainee.

- Name the data sets **`US_trainees`** and **`AU_trainees`**. For this exercise, a trainee is anyone that has the job title of Sales Rep. I
- Each data set should contain the fields indicated by arrows in the layout table.
- Write only U.S. trainees to the **`US_trainees`** data set and only Australian trainees to the **`AU_trainees`** data set. Do **not** keep the **`Country`** variable in the output data sets.

b. Print both of the data sets with appropriate titles.

Partial PROC PRINT Output for **`AU_trainees`** (21 Total Observations)

Australian Trainees

Employee_ ID	Last_Name	Job_Title	Salary	Hire_ Date
120123	Hotstone	Sales Rep. I	26190	17440
120124	Daymond	Sales Rep. I	26480	17226
120130	Lyon	Sales Rep. I	26955	17287
120131	Surawski	Sales Rep. I	26910	17167
120136	Leyden	Sales Rep. I	26605	17198
120138	Duckett	Sales Rep. I	25795	17348

Partial PROC PRINT Output for **`US_trainees`** (42 Total Observations)

US Trainees

Employee_ ID	Last_Name	Job_Title	Salary	Hire_ Date
121023	Fuller	Sales Rep. I	26010	17287
121028	Smades	Sales Rep. I	26585	17471
121029	Mcelwee	Sales Rep. I	27225	17501
121030	Areu	Sales Rep. I	26745	17198
121036	Mesley	Sales Rep. I	25965	17440
121038	Anstey	Sales Rep. I	25285	17379
121044	Abbott	Sales Rep. I	25660	17379

3. Using a Text String with Column Pointer Controls

- The raw data file **`seminar.dat`** contains comments and ratings from participants at a seminar given to Orion Star sales staff.
- The data file contains one line for each participant:
  - The first 15 characters are reserved for the name of the participant (if given).
  - There can be a comment of up to 60 characters.
  - There will be the text Rating: immediately followed by a numeric score from 1 to 5.

Listing of **`seminar.dat`**

```
J. Mitchell         Very Well done!  Rating:5
Amy Jung            Rating:4
Carl Heisman        Rating:4
Linda Deal          Not enough give aways  Rating:3
Gabrielle Heron     Nice! Rating:4
                    Not helpful at all Rating:2
Kyle Patterson      Very good. Need more like it  Rating:5
```

a. Create a new SAS data set named **`seminar_ratings`** that contains the names of the participants and the ratings that were given.

b. Print the data set and give it an appropriate title.

PROC PRINT Output

Names and Ratings

Obs	Name	Rating
1	J. Mitchell	5
2	Amy Jung	4
3	Carl Heisman	4
4	Linda Deal	3
5	Gabrielle Heron	4
6		2
7	Kyle Patterson	5

4. Reading Multiple Input Records per Observation

- The raw data file **`sales2.dat`** has employee information for the Australian and U.S. sales staff.
- Information for each employee is in three lines of raw data.
- The record layouts are shown below.

Line 1 layout

	Field Description	Starting Column	Length of Field	Data Type
➡	Employee ID	1	6	Numeric
	First Name	8	12	Character
➡	Last Name	21	18	Character

Line 2 layout

	Field Description	Starting Column	Length of Field	Data Type
➡	Job Title	1	20	Character
➡	Hire Date	22	10	Numeric mm/dd/yyyy
➡	Salary	33	8	Numeric for example, $100,000

Line 3 layout

Field Description	Starting Column	Length of Field	Data Type
Gender	1	1	Character
Birth Date	3	10	Numeric mm/dd/yyyy
Country	14	2	Character

a. Create a new SAS data set named **`sales_staff2`** that contains the fields indicated by arrows in the layout table.

b. Print **`sales_staff2`** and add an appropriate title.

Partial PROC PRINT Output (165 Total Observations)

Australian and US Sales Staff

Employee_ ID	Last_Name	Job_Title	Hire_ Date	Salary
120102	Zhou	Sales Manager	10744	108255
120103	Dawes	Sales Manager	5114	87975
120121	Elvish	Sales Rep. II	5114	26600
120122	Ngan	Sales Rep. II	6756	27475
120123	Hotstone	Sales Rep. I	17440	26190
120124	Daymond	Sales Rep. I	17226	26480

5. Working with Mixed Record Types

- The raw data file **`sales3.dat`** has employee information for the Australian and U.S. Ssales staff.
- Information for each employee is in two lines of raw data.
- The record layouts are shown below.

Line 1 layout

Field Description	Starting Column	Length of Field	Data Type
Employee ID	1	6	Numeric
First Name	8	12	Character
Last Name	21	18	Character
Gender	40	1	Character
Job Title	43	20	Character

Line 2 layout for Australian employees

Field Description	Starting Column	Length of Field	Data Type
Salary	1	8	Numeric $100,000
Country	10	2	Character
Birth Date	13	10	Numeric dd/mm/yyyy
Hire Date	24	10	Numeric dd/mm/yyyy

Line 2 layout for U.S. employees

	Field Description	Starting Column	Length of Field	Data Type
➡	Salary	1	8	Numeric $100,000
➡	Country	10	2	Character
	Birth Date	13	10	Numeric mm/dd/yyyy
➡	Hire Date	24	10	Numeric mm/dd/yyyy

a. Create two new SAS data sets, **`US_sales`** and **`AU_sales`**, that contain the fields indicated by arrows in the layout table. Write only U.S. employees to the **`US_sales`** data set and only Australian employees to the **`AU_sales`** data set. Do **not** include the **`Country`** variable in the output data sets.

The salary and hire date values are different for Australian and U.S. employees. Be sure to use the correct informats in each INPUT statement.

b. Print both of the data sets with appropriate titles.

Partial PROC PRINT Output for **`AU_sales`** (63 Total Observations)

Australian Sales Staff

Employee_ ID	Last_Name	Job_Title	Salary	Hire_ Date
120102	Zhou	Sales Manager	108255	10744
120103	Dawes	Sales Manager	87975	5114
120121	Elvish	Sales Rep. II	26600	5114
120122	Ngan	Sales Rep. II	27475	6756
120123	Hotstone	Sales Rep. I	26190	17440
120124	Daymond	Sales Rep. I	26480	17226
120125	Hofmeister	Sales Rep. IV	32040	6999

Partial PROC PRINT Output for **`US_sales`** (102 Total Observations)

US Sales Staff

Employee_ ID	Last_Name	Job_Title	Salary	Hire_ Date
120261	Highpoint	Chief Sales Officer	243190	10074
121018	Magolan	Sales Rep. II	27560	5114
121019	Desanctis	Sales Rep. IV	31320	16223
121020	Ridley	Sales Rep. IV	31750	15461
121021	Farren	Sales Rep. IV	32985	12478
121022	Stevens	Sales Rep. IV	32210	15372
121023	Fuller	Sales Rep. I	26010	17287
121024	Westlund	Sales Rep. II	26600	16192