**Q1:**

* **1)**

FILENAME tennis "/folders/myfolders/sasuser.v94/tennis for hw2.dat";  
DATA tennis;  
INFILE tennis firstobs=3;  
INPUT name $ 1-22 matches wins losses bagels\_served bagels\_eaten male;  
RUN;  
PROC PRINT DATA=tennis;  
RUN;

* **2)**

FILENAME tennis1 "/folders/myfolders/sasuser.v94/tennis for hw2.dat";

DATA tennis1;

INFILE tennis1 firstobs=3;

INPUT name $ 1-22 matches wins losses male;

RUN;

PROC PRINT DATA=tennis1;

RUN;

RESULT:

| **Obs** | **name** | **matches** | **wins** | **losses** | **male** |
| --- | --- | --- | --- | --- | --- |
| **1** | Roger Federer | 64 | 54 | 10 | 1 |
| **2** | Marin Cilic | 61 | 45 | 16 | 4 |
| **3** | David Ferrer | 61 | 44 | 17 | 8 |
| **4** | Fabio Fognini | 58 | 39 | 19 | 2 |
| **5** | Tomas Berdych | 56 | 41 | 15 | 2 |
| **6** | Roberto Bautista Agut | 54 | 36 | 18 | 1 |
| **7** | Feliciano Lopez | 54 | 33 | 21 | 0 |
| **8** | Philipp Kohlschreiber | 54 | 33 | 21 | 0 |
| **9** | Grigor Dimitrov | 53 | 40 | 13 | 0 |
| **10** | Andy Murray | 53 | 38 | 15 | 5 |
| **11** | Rafael Nadal | 52 | 44 | 8 | 6 |
| **12** | Ernests Gulbis | 52 | 36 | 16 | 3 |
| **13** | Kevin Anderson | 52 | 33 | 19 | 1 |
| **14** | Tommy Robredo | 52 | 32 | 20 | 1 |
| **15** | Ivo Karlovic | 52 | 29 | 23 | 0 |
| **16** | Novak Djokovic | 51 | 44 | 7 | 6 |
| **17** | Kei Nishikori | 50 | 40 | 10 | 2 |
| **18** | Milos Raonic | 50 | 37 | 13 | 1 |
| **19** | John Isner | 49 | 33 | 16 | 0 |
| **20** | Lukas Rosol | 49 | 28 | 21 | 0 |
| **21** | Ana Ivanovic | 64 | 49 | 15 | 4 |
| **22** | Carla Suarez Navarro | 64 | 44 | 20 | 5 |
| **23** | Angelique Kerber | 60 | 41 | 19 | 4 |
| **24** | Agnieszka Radwanska | 57 | 41 | 16 | 1 |
| **25** | Simona Halep | 55 | 41 | 14 | 1 |
| **26** | Caroline Wozniacki | 55 | 40 | 15 | 1 |
| **27** | Sara Errani | 55 | 34 | 21 | 6 |
| **28** | Ekaterina Makarova | 54 | 38 | 16 | 3 |
| **29** | Jelena Jankovic | 54 | 36 | 18 | 5 |
| **30** | Eugenie Bouchard | 53 | 36 | 17 | 5 |
| **31** | Serena Williams | 51 | 45 | 6 | 6 |
| **32** | Maria Sharapova | 51 | 41 | 10 | 8 |
| **33** | Andrea Petkovic | 51 | 34 | 17 | 4 |
| **34** | Lucie Safarova | 51 | 32 | 19 | 5 |
| **35** | Klara Koukalova | 51 | 28 | 23 | 4 |
| **36** | Garbine Muguruza | 48 | 30 | 18 | 2 |
| **37** | Dominika Cibulkova | 48 | 29 | 19 | 5 |
| **38** | Elina Svitolina | 48 | 28 | 20 | 1 |
| **39** | Karolina Pliskova | 46 | 28 | 18 | 3 |
| **40** | Shuai Peng | 46 | 28 | 18 | 3 |

* **3)**

FILENAME tennis1 "/folders/myfolders/sasuser.v94/tennis for hw2.dat";

DATA tennis1;

INFILE tennis1 firstobs=3;

INPUT name $ 1-22 matches wins losses male;

RUN;

DATA tennis2;

SET tennis1;

pct = wins/matches;

if male=1 then tour = "ATP";

if male=0 then tour = "WTP";

PROC print data=tennis2;

RUN;

RESULT:

| **Obs** | **name** | **matches** | **wins** | **losses** | **male** | **pct** | **tour** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | Roger Federer | 64 | 54 | 10 | 1 | 0.84375 | ATP |
| **2** | Marin Cilic | 61 | 45 | 16 | 4 | 0.73770 |  |
| **3** | David Ferrer | 61 | 44 | 17 | 8 | 0.72131 |  |
| **4** | Fabio Fognini | 58 | 39 | 19 | 2 | 0.67241 |  |
| **5** | Tomas Berdych | 56 | 41 | 15 | 2 | 0.73214 |  |
| **6** | Roberto Bautista Agut | 54 | 36 | 18 | 1 | 0.66667 | ATP |
| **7** | Feliciano Lopez | 54 | 33 | 21 | 0 | 0.61111 | WTP |
| **8** | Philipp Kohlschreiber | 54 | 33 | 21 | 0 | 0.61111 | WTP |
| **9** | Grigor Dimitrov | 53 | 40 | 13 | 0 | 0.75472 | WTP |
| **10** | Andy Murray | 53 | 38 | 15 | 5 | 0.71698 |  |
| **11** | Rafael Nadal | 52 | 44 | 8 | 6 | 0.84615 |  |
| **12** | Ernests Gulbis | 52 | 36 | 16 | 3 | 0.69231 |  |
| **13** | Kevin Anderson | 52 | 33 | 19 | 1 | 0.63462 | ATP |
| **14** | Tommy Robredo | 52 | 32 | 20 | 1 | 0.61538 | ATP |
| **15** | Ivo Karlovic | 52 | 29 | 23 | 0 | 0.55769 | WTP |
| **16** | Novak Djokovic | 51 | 44 | 7 | 6 | 0.86275 |  |
| **17** | Kei Nishikori | 50 | 40 | 10 | 2 | 0.80000 |  |
| **18** | Milos Raonic | 50 | 37 | 13 | 1 | 0.74000 | ATP |
| **19** | John Isner | 49 | 33 | 16 | 0 | 0.67347 | WTP |
| **20** | Lukas Rosol | 49 | 28 | 21 | 0 | 0.57143 | WTP |
| **21** | Ana Ivanovic | 64 | 49 | 15 | 4 | 0.76563 |  |
| **22** | Carla Suarez Navarro | 64 | 44 | 20 | 5 | 0.68750 |  |
| **23** | Angelique Kerber | 60 | 41 | 19 | 4 | 0.68333 |  |
| **24** | Agnieszka Radwanska | 57 | 41 | 16 | 1 | 0.71930 | ATP |
| **25** | Simona Halep | 55 | 41 | 14 | 1 | 0.74545 | ATP |
| **26** | Caroline Wozniacki | 55 | 40 | 15 | 1 | 0.72727 | ATP |
| **27** | Sara Errani | 55 | 34 | 21 | 6 | 0.61818 |  |
| **28** | Ekaterina Makarova | 54 | 38 | 16 | 3 | 0.70370 |  |
| **29** | Jelena Jankovic | 54 | 36 | 18 | 5 | 0.66667 |  |
| **30** | Eugenie Bouchard | 53 | 36 | 17 | 5 | 0.67925 |  |
| **31** | Serena Williams | 51 | 45 | 6 | 6 | 0.88235 |  |
| **32** | Maria Sharapova | 51 | 41 | 10 | 8 | 0.80392 |  |
| **33** | Andrea Petkovic | 51 | 34 | 17 | 4 | 0.66667 |  |
| **34** | Lucie Safarova | 51 | 32 | 19 | 5 | 0.62745 |  |
| **35** | Klara Koukalova | 51 | 28 | 23 | 4 | 0.54902 |  |
| **36** | Garbine Muguruza | 48 | 30 | 18 | 2 | 0.62500 |  |
| **37** | Dominika Cibulkova | 48 | 29 | 19 | 5 | 0.60417 |  |
| **38** | Elina Svitolina | 48 | 28 | 20 | 1 | 0.58333 | ATP |
| **39** | Karolina Pliskova | 46 | 28 | 18 | 3 | 0.60870 |  |
| **40** | Shuai Peng | 46 | 28 | 18 | 3 | 0.60870 |  |

* **4)**

FILENAME tennis1 "/folders/myfolders/sasuser.v94/tennis for hw2.dat";  
 DATA tennis1;  
 INFILE tennis1 firstobs=3;  
 INPUT name $ 1-22 matches wins losses male;  
RUN;  
  
DATA tennis2;

SET tennis1;

pct = wins/matches;

if male=1 then tour = "ATP";

if male=0 then tour = "WTP";  
  
PROC SORT DATA=tennis2 OUT=tennis3 NODUPKEY;

BY pct DESCENDING pct;  
RUN;   
  
PROC SORT DATA=tennis2 OUT=tennis3 NODUPKEY;

BY pct DESCENDING pct;  
RUN;   
  
PROC PRINT DATA=tennis3;  
RUN;

RESULT:

| **Obs** | **name** | **matches** | **wins** | **losses** | **male** | **pct** | **tour** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | Klara Koukalova | 51 | 28 | 23 | 4 | 0.54902 |  |
| **2** | Ivo Karlovic | 52 | 29 | 23 | 0 | 0.55769 | WTP |
| **3** | Lukas Rosol | 49 | 28 | 21 | 0 | 0.57143 | WTP |
| **4** | Elina Svitolina | 48 | 28 | 20 | 1 | 0.58333 | ATP |
| **5** | Dominika Cibulkova | 48 | 29 | 19 | 5 | 0.60417 |  |
| **6** | Karolina Pliskova | 46 | 28 | 18 | 3 | 0.60870 |  |
| **7** | Feliciano Lopez | 54 | 33 | 21 | 0 | 0.61111 | WTP |
| **8** | Tommy Robredo | 52 | 32 | 20 | 1 | 0.61538 | ATP |
| **9** | Sara Errani | 55 | 34 | 21 | 6 | 0.61818 |  |
| **10** | Garbine Muguruza | 48 | 30 | 18 | 2 | 0.62500 |  |
| **11** | Lucie Safarova | 51 | 32 | 19 | 5 | 0.62745 |  |
| **12** | Kevin Anderson | 52 | 33 | 19 | 1 | 0.63462 | ATP |
| **13** | Roberto Bautista Agut | 54 | 36 | 18 | 1 | 0.66667 | ATP |
| **14** | Fabio Fognini | 58 | 39 | 19 | 2 | 0.67241 |  |
| **15** | John Isner | 49 | 33 | 16 | 0 | 0.67347 | WTP |
| **16** | Eugenie Bouchard | 53 | 36 | 17 | 5 | 0.67925 |  |
| **17** | Angelique Kerber | 60 | 41 | 19 | 4 | 0.68333 |  |
| **18** | Carla Suarez Navarro | 64 | 44 | 20 | 5 | 0.68750 |  |
| **19** | Ernests Gulbis | 52 | 36 | 16 | 3 | 0.69231 |  |
| **20** | Ekaterina Makarova | 54 | 38 | 16 | 3 | 0.70370 |  |
| **21** | Andy Murray | 53 | 38 | 15 | 5 | 0.71698 |  |
| **22** | Agnieszka Radwanska | 57 | 41 | 16 | 1 | 0.71930 | ATP |
| **23** | David Ferrer | 61 | 44 | 17 | 8 | 0.72131 |  |
| **24** | Caroline Wozniacki | 55 | 40 | 15 | 1 | 0.72727 | ATP |
| **25** | Tomas Berdych | 56 | 41 | 15 | 2 | 0.73214 |  |
| **26** | Marin Cilic | 61 | 45 | 16 | 4 | 0.73770 |  |
| **27** | Milos Raonic | 50 | 37 | 13 | 1 | 0.74000 | ATP |
| **28** | Simona Halep | 55 | 41 | 14 | 1 | 0.74545 | ATP |
| **29** | Grigor Dimitrov | 53 | 40 | 13 | 0 | 0.75472 | WTP |
| **30** | Ana Ivanovic | 64 | 49 | 15 | 4 | 0.76563 |  |
| **31** | Kei Nishikori | 50 | 40 | 10 | 2 | 0.80000 |  |
| **32** | Maria Sharapova | 51 | 41 | 10 | 8 | 0.80392 |  |
| **33** | Roger Federer | 64 | 54 | 10 | 1 | 0.84375 | ATP |
| **34** | Rafael Nadal | 52 | 44 | 8 | 6 | 0.84615 |  |
| **35** | Novak Djokovic | 51 | 44 | 7 | 6 | 0.86275 |  |
| **36** | Serena Williams | 51 | 45 | 6 | 6 | 0.88235 |  |

* **5)**

FILENAME tennis1 "/folders/myfolders/sasuser.v94/tennis for hw2.dat";  
 DATA tennis1;  
 INFILE tennis1 firstobs=3;  
 INPUT name $ 1-22 matches wins losses male;  
RUN;  
  
DATA tennis2;

SET tennis1;

pct = wins/matches;

if male=1 then tour = "ATP";

if male=0 then tour = "WTP";  
  
PROC SORT DATA=tennis2 OUT=tennis3 NODUPKEY;

BY pct DESCENDING pct;  
RUN;   
  
DATA atpdata;  
 SET tennis3;

if male=1;  
PROC PRINT DATA=atpdata;  
RUN;   
  
DATA wtadata;  
 SET tennis3;

if male=0;  
PROC PRINT DATA=wtadata;  
RUN;

RESULTS:

| **Obs** | **name** | **matches** | **wins** | **losses** | **male** | **pct** | **tour** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | Elina Svitolina | 48 | 28 | 20 | 1 | 0.58333 | ATP |
| **2** | Tommy Robredo | 52 | 32 | 20 | 1 | 0.61538 | ATP |
| **3** | Kevin Anderson | 52 | 33 | 19 | 1 | 0.63462 | ATP |
| **4** | Roberto Bautista Agut | 54 | 36 | 18 | 1 | 0.66667 | ATP |
| **5** | Agnieszka Radwanska | 57 | 41 | 16 | 1 | 0.71930 | ATP |
| **6** | Caroline Wozniacki | 55 | 40 | 15 | 1 | 0.72727 | ATP |
| **7** | Milos Raonic | 50 | 37 | 13 | 1 | 0.74000 | ATP |
| **8** | Simona Halep | 55 | 41 | 14 | 1 | 0.74545 | ATP |
| **9** | Roger Federer | 64 | 54 | 10 | 1 | 0.84375 | ATP |

| **Obs** | **name** | **matches** | **wins** | **losses** | **male** | **pct** | **tour** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | Ivo Karlovic | 52 | 29 | 23 | 0 | 0.55769 | WTP |
| **2** | Lukas Rosol | 49 | 28 | 21 | 0 | 0.57143 | WTP |
| **3** | Philipp Kohlschreiber | 54 | 33 | 21 | 0 | 0.61111 | WTP |
| **4** | Feliciano Lopez | 54 | 33 | 21 | 0 | 0.61111 | WTP |
| **5** | John Isner | 49 | 33 | 16 | 0 | 0.67347 | WTP |
| **6** | Grigor Dimitrov | 53 | 40 | 13 | 0 | 0.75472 | WTP |

**6)**

FILENAME tennis1 "/folders/myfolders/sasuser.v94/tennis for hw2.dat";  
 DATA tennis1;  
 INFILE tennis1 firstobs=3;  
 INPUT name $ 1-22 matches wins losses male;  
RUN;  
  
DATA tennis2;

SET tennis1;

pct = wins/matches;

if male=1 then tour = "ATP";

if male=0 then tour = "WTP";  
  
PROC SORT DATA=tennis2 OUT=tennis3 NODUPKEY;

BY pct DESCENDING pct;  
RUN;   
  
DATA atpdata;  
 SET tennis3;

if male=1;  
RUN;   
  
data \_null\_;  
 set atpdata;  
 file "/folders/myfolders/sasuser.v94/atp\_rankings.txt";  
 put name pct;  
 file print;  
 put name pct;run;

RESULTS:

|  |
| --- |
| Elina Svitolina 0.5833333333 |
| Tommy Robredo 0.6153846154 |
| Kevin Anderson 0.6346153846 |
| Roberto Bautista Agut 0.6666666667 |
| Agnieszka Radwanska 0.7192982456 |
| Caroline Wozniacki 0.7272727273 |
| Milos Raonic 0.74 |
| Simona Halep 0.7454545455 |
| Roger Federer 0.84375 |

**Q2:**

* **1)**

These methods do not require sorting: Concatenating, One to one merge

1. **Concatenating:**

DATA One;  
INPUT ID age name $;  
DATALINES;  
1001 21 Jone

1002 20 Peter

1003 22 Mary

1004 19 Joe

1005 21 Mark

1006 23 Sue

1007 19 Harry

1008 18 Tom

1009 22 Andy

1020 21 Larry  
;  
RUN;  
  
DATA Two;  
INPUT ID gender $ department $;  
DATALINES;  
1006 F Econ

1007 M Econ

1008 M Math

1009 M Stat

1010 M Stat

1001 M Econ

1002 M Stat

1003 F Econ

1004 F Econ

1005 M Fina  
;  
RUN;  
  
DATA Concatenating;  
 SET One Two;  
RUN;  
PROC PRINT data=Concatenating;  
RUN;

RESULT:

| **Obs** | **ID** | **age** | **name** | **gender** | **department** |
| --- | --- | --- | --- | --- | --- |
| **1** | 1001 | 21 | Jone |  |  |
| **2** | 1002 | 20 | Peter |  |  |
| **3** | 1003 | 22 | Mary |  |  |
| **4** | 1004 | 19 | Joe |  |  |
| **5** | 1005 | 21 | Mark |  |  |
| **6** | 1006 | 23 | Sue |  |  |
| **7** | 1007 | 19 | Harry |  |  |
| **8** | 1008 | 18 | Tom |  |  |
| **9** | 1009 | 22 | Andy |  |  |
| **10** | 1020 | 21 | Larry |  |  |
| **11** | 1006 | . |  | F | Econ |
| **12** | 1007 | . |  | M | Econ |
| **13** | 1008 | . |  | M | Math |
| **14** | 1009 | . |  | M | Stat |
| **15** | 1010 | . |  | M | Stat |
| **16** | 1001 | . |  | M | Econ |
| **17** | 1002 | . |  | M | Stat |
| **18** | 1003 | . |  | F | Econ |
| **19** | 1004 | . |  | F | Econ |
| **20** | 1005 | . |  | M | Fina |

**b) One to one merge:**

DATA One;  
INPUT ID age name $;  
DATALINES;

1001 21 Jone

1002 20 Peter

1003 22 Mary

1004 19 Joe

1005 21 Mark

1006 23 Sue

1007 19 Harry

1008 18 Tom

1009 22 Andy

1020 21 Larry

;  
RUN;  
  
DATA Two;  
INPUT ID gender $ department $;  
DATALINES;  
1006 F Econ

1007 M Econ

1008 M Math

1009 M Stat

1010 M Stat

1001 M Econ

1002 M Stat

1003 F Econ

1004 F Econ

1005 M Fina  
;  
RUN;  
  
DATA One\_to\_one\_merge;  
 Merge One Two;  
RUN;  
PROC PRINT data=One\_to\_one\_merge;  
RUN;

RESULT:

| **Obs** | **ID** | **age** | **name** | **gender** | **department** |
| --- | --- | --- | --- | --- | --- |
| **1** | 1006 | 21 | Jone | F | Econ |
| **2** | 1007 | 20 | Peter | M | Econ |
| **3** | 1008 | 22 | Mary | M | Math |
| **4** | 1009 | 19 | Joe | M | Stat |
| **5** | 1010 | 21 | Mark | M | Stat |
| **6** | 1001 | 23 | Sue | M | Econ |
| **7** | 1002 | 19 | Harry | M | Stat |
| **8** | 1003 | 18 | Tom | F | Econ |
| **9** | 1004 | 22 | Andy | F | Econ |
| **10** | 1005 | 21 | Larry | M | Fina |

* **2)**

These methods require sorting: Interleaving, Match merge

1. **Interleaving:**

DATA One;  
INPUT ID age name $;  
DATALINES;

1001 21 Jone

1002 20 Peter

1003 22 Mary

1004 19 Joe

1005 21 Mark

1006 23 Sue

1007 19 Harry

1008 18 Tom

1009 22 Andy

1020 21 Larry  
;  
RUN;  
  
DATA Two;  
INPUT ID gender $ department $;  
DATALINES;

1006 F Econ

1007 M Econ

1008 M Math

1009 M Stat

1010 M Stat

1001 M Econ

1002 M Stat

1003 F Econ

1004 F Econ

1005 M Fina  
;  
RUN;  
  
PROC SORT data=One;  
 BY ID; RUN;  
PROC SORT data=Two;  
 BY Id; RUN;  
DATA Interleaving;  
 SET One Two;  
 BY ID; RUN;  
PROC PRINT data=Interleaving; RUN;

RESULT:

| **Obs** | **ID** | **age** | **name** | **gender** | **department** |
| --- | --- | --- | --- | --- | --- |
| **1** | 1001 | 21 | Jone |  |  |
| **2** | 1001 | . |  | M | Econ |
| **3** | 1002 | 20 | Peter |  |  |
| **4** | 1002 | . |  | M | Stat |
| **5** | 1003 | 22 | Mary |  |  |
| **6** | 1003 | . |  | F | Econ |
| **7** | 1004 | 19 | Joe |  |  |
| **8** | 1004 | . |  | F | Econ |
| **9** | 1005 | 21 | Mark |  |  |
| **10** | 1005 | . |  | M | Fina |
| **11** | 1006 | 23 | Sue |  |  |
| **12** | 1006 | . |  | F | Econ |
| **13** | 1007 | 19 | Harry |  |  |
| **14** | 1007 | . |  | M | Econ |
| **15** | 1008 | 18 | Tom |  |  |
| **16** | 1008 | . |  | M | Math |
| **17** | 1009 | 22 | Andy |  |  |
| **18** | 1009 | . |  | M | Stat |
| **19** | 1010 | . |  | M | Stat |
| **20** | 1020 | 21 | Larry |  |  |

**b) Match Merge:**

DATA One;

INPUT ID age name $;

DATALINES;

1001 21 Jone

1002 20 Peter

1003 22 Mary

1004 19 Joe

1005 21 Mark

1006 23 Sue

1007 19 Harry

1008 18 Tom

1009 22 Andy

1020 21 Larry

;

RUN;

PROC SORT data=One;

BY ID; RUN;

DATA Two;

INPUT ID gender $ department $;

DATALINES;

1006 F Econ

1007 M Econ

1008 M Math

1009 M Stat

1010 M Stat

1001 M Econ

1002 M Stat

1003 F Econ

1004 F Econ

1005 M Fina

;

RUN;

PROC SORT data=Two;

BY ID; RUN;

DATA Match\_merge;

MERGE One Two;

BY ID;

RUN;

PROC PRINT data=Match\_merge;

RUN;

RESULT:

| **Obs** | **ID** | **age** | **name** | **gender** | **department** |
| --- | --- | --- | --- | --- | --- |
| **1** | 1001 | 21 | Jone | M | Econ |
| **2** | 1002 | 20 | Peter | M | Stat |
| **3** | 1003 | 22 | Mary | F | Econ |
| **4** | 1004 | 19 | Joe | F | Econ |
| **5** | 1005 | 21 | Mark | M | Fina |
| **6** | 1006 | 23 | Sue | F | Econ |
| **7** | 1007 | 19 | Harry | M | Econ |
| **8** | 1008 | 18 | Tom | M | Math |
| **9** | 1009 | 22 | Andy | M | Stat |
| **10** | 1010 | . |  | M | Stat |
| **11** | 1020 | 21 | Larry |  |  |