

# PDSP 2022, Practice Assignment 1, 09 August 2022

## Placement data as a list

All questions below use the same list of placement data that we had seen in the examples in Lecture 2, 4 August 2022.

```
In [1]: plist = [
("2017-18", "B.Sc", "CS", 1800000),
("2017-18", "M.Sc Appl. Math", "CS", 1170000),
("2017-18", "M.Sc Appl. Math", "Manufacturing", 730000),
("2017-18", "M.Sc Appl. Math", "Banking-Finance", 1490000),
("2017-18", "M.Sc Appl. Math", "Banking-Finance", 1000000),
("2017-18", "M.Sc Appl. Math", "Logistics", 1250000),
("2017-18", "M.Sc Appl. Math", "Logistics", 1250000),
("2017-18", "M.Sc Appl. Math", "Banking-Finance", 1155000),
("2017-18", "M.Sc Appl. Math", "Banking-Finance", 1155000),
("2017-18", "M.Sc Appl. Math", "CS", 2000000),
("2018-19", "B.Sc", "Banking-Finance", 700000),
("2018-19", "B.Sc", "Banking-Finance", 1480000),
("2018-19", "M.Sc Appl. Math", "Manufacturing", 730000),
("2019-20", "M.Sc Data Science", "CS", 2000000),
("2019-20", "M.Sc Data Science", "CS", 1800000),
("2019-20", "M.Sc Data Science", "Analytics", 1700000),
("2019-20", "M.Sc Data Science", "Analytics", 1700000),
("2019-20", "M.Sc Data Science", "Analytics", 1350000),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1344000),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Comp. Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Comp. Science", "Banking-Finance", 1335827),
("2019-20", "M.Sc Data Science", "CS", 1301968),
("2019-20", "M.Sc Data Science", "CS", 1301968),
("2019-20", "M.Sc Comp. Science", "Logistics", 1000000),
("2019-20", "M.Sc Data Science", "Banking-Finance", 1600000),
("2020-21", "M.Sc Comp. Science", "Banking-Finance", 1500000),
("2020-21", "M.Sc Data Science", "Analytics", 900000),
("2020-21", "Ph.D Comp. Science", "Banking-Finance", 1454545),
("2020-21", "M.Sc Data Science", "Analytics", 1300000),
("2020-21", "M.Sc Data Science", "Analytics", 700000),
("2020-21", "M.Sc Data Science", "Analytics", 700000),
("2020-21", "M.Sc Data Science", "Analytics", 700000),
("2020-21", "M.Sc Data Science", "Analytics", 700000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1350000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1350000),
("2020-21", "M.Sc Data Science", "Analytics", 1680000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1360000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1300000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1300000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 900000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1842632),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1842632),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1842632),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1842632),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1842632),
("2020-21", "M.Sc Data Science", "CS", 1150000),
("2020-21", "M.Sc Data Science", "Analytics", 850000),
("2020-21", "M.Sc Data Science", "Analytics", 850000),
("2020-21", "M.Sc Comp. Science", "Analytics", 850000),
("2020-21", "M.Sc Data Science", "CS", 1200000),
("2020-21", "M.Sc Data Science", "Analytics", 1608000),
("2020-21", "M.Sc Data Science", "Analytics", 1608000),
("2020-21", "M.Sc Data Science", "CS", 1100000),
("2020-21", "M.Sc Data Science", "Analytics", 1500000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1550000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1550000),
("2020-21", "M.Sc Comp. Science", "Banking-Finance", 1350000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1200000),
("2020-21", "M.Sc Data Science", "Analytics", 900000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 3300000),
("2020-21", "Ph.D Physics", "Banking-Finance", 3000000),
("2020-21", "M.Sc Data Science", "Banking-Finance", 1500000),
]
```

## Questions

1. Compute the minimum pay package overall.

```
In [5]: min = plist[0][3]
        for row in plist:
            if row[3] < min:
                min = row[3]
```

```
In [6]: min
```

```
Out[6]: 700000
```

2. Compute the second lowest pay package overall. There may be multiple students with the same minimum pay package. The value we want is the next lowest pay package other than the minimum pay package.

```
In [14]: sec_min = plist[0][3]
         for row in plist:
             if row[3] < sec_min and row[3] > min:
                 sec_min = row[3]
```

```
In [15]: sec_min
```

```
Out[15]: 730000
```

3. Compute the average (mean) pay package of students graduating with B.Sc.

```
In [17]: bsc_count = 0
         bsc_sum = 0
         for row in plist:
             if row[1] == "B.Sc":
                 bsc_count = bsc_count + 1
                 bsc_sum = bsc_sum + row[3]
```

```
In [19]: bsc_avg = bsc_sum/bsc_count
         bsc_avg
```

```
Out[19]: 1326666.6666666667
```

4. Compute whether the following statement is true: the average pay package for M.Sc Comp. Science is lower than the average pay package for M.Sc Data Science.

```
In [26]: sum_count_dict = {} # key = stream, value = (sum,count)
         for row in plist:
             stream = row[1]
             if stream in sum_count_dict:
                 sum_count_dict[stream][0] = sum_count_dict[stream][0] + row[3]
                 sum_count_dict[stream][1] = sum_count_dict[stream][1] + 1
             else:
                 sum_count_dict[stream] = [row[3],1]
```

```
In [28]: sum_count_dict
```

```
Out[28]: {'B.Sc': [3980000, 3],
          'M.Sc Appl. Math': [11930000, 10],
          'M.Sc Data Science': [72111193, 52],
          'M.Sc Comp. Science': [7371654, 6],
          'Ph.D Comp. Science': [1454545, 1],
          'Ph.D Physics': [3000000, 1]}
```

```
In [34]: msc_cs_mean = sum_count_dict['M.Sc Comp. Science'][0]/sum_count_dict['M.Sc Comp. Science'][1]
         msc_ds_mean = sum_count_dict['M.Sc Data Science'][0]/ sum_count_dict['M.Sc Data Science'][1]

         print(msc_cs_mean,msc_ds_mean)
```

```
1228609.0 1386753.7115384615
```

5. Compute the graduating degree (B.Sc, M.Sc Comp. Science, M.Sc Data Science, ...) with the minimum average pay package.

```
In [46]: avg_dict = {}
for key in sum_count_dict:
    avg_dict[key] = sum_count_dict[key][0]/sum_count_dict[key][1]
```

```
In [47]: # list(avg_dict.keys())[0]
```

```
In [38]: min_stream = 'B.Sc'
min_pay = avg_dict['B.Sc']
for key in avg_dict:
    if avg_dict[key] < min_pay:
        min_stream = key
        min_pay = avg_dict[key]
```

```
In [39]: min_stream, min_pay
```

```
Out[39]: ('M.Sc Appl. Math', 1193000.0)
```

6. Compute the domain (Analytics, Banking-Finance, ...) in which the maximum number of different graduating degrees (B.Sc, M.Sc Comp. Science, M.Sc Data Science, ...) have been placed. For instance, in Analytics, students from M.Sc Comp. Science and M.Sc Data Science have been placed.

```
In [49]: # {
#   'Analytics' : ['Bsc-Math', 'Msc-Math', 'Phd-CS']
#   'IT' : ['Msc-Cs', 'Msc-Ds']
#   'Finance' : ['Phd-Physics', 'BSc-math']
#   'Medical' : ['PhD-Comp-Bio']
# }

domain_stream_dict = {}
for row in plist:
    domain = row[2]
    stream = row[1]
    if domain not in domain_stream_dict:
        domain_stream_dict[domain] = [stream]
    else:
        if stream in domain_stream_dict[domain]:
            pass
        else:
            domain_stream_dict[domain].append(stream)
```

```
In [50]: domain_stream_dict
```

```
Out[50]: {'CS': ['B.Sc', 'M.Sc Appl. Math', 'M.Sc Data Science'],
'Manufacturing': ['M.Sc Appl. Math'],
'Banking-Finance': ['M.Sc Appl. Math',
'B.Sc',
'M.Sc Data Science',
'M.Sc Comp. Science',
'Ph.D Comp. Science',
'Ph.D Physics'],
'Logistics': ['M.Sc Appl. Math', 'M.Sc Comp. Science'],
'Analytics': ['M.Sc Data Science', 'M.Sc Comp. Science']}
```

```
In [51]: domain_name = plist[0][2]
domain_freq = 0

for key in domain_stream_dict:
    count = len(domain_stream_dict[key])
    if count > domain_freq:
        domain_freq = count
        domain_name = key
```

```
In [52]: domain_name, domain_freq
```

```
Out[52]: ('Banking-Finance', 6)
```

```
In [ ]:
```