TITLE :

Analyze Titanic Disaster Data Using MapReduce
I. Calculate the average age of deceased males.
II. Determine the number of deceased females in each class.

NAME : Shinde Shubham Dnyandev.   **ROLL NO :** 23107121.   **CLASS :** TY-B

PROGRAM :

MAPPER :

```python
#!/usr/bin/env python3
# mapper.py
# Expects CSV columns like:
# PassengerId,Survived,Pclass,Name,Sex,Age,...  (standard Titanic layout)
import sys, csv

def safe_strip(s):
    return s.strip()

reader = csv.reader(sys.stdin)
header = next(reader, None)

# try to discover column indices if header present
if header:
    try:
        idx_survived = header.index('Survived')
        idx_pclass = header.index('Pclass')
        idx_sex = header.index('Sex')
        idx_age = header.index('Age')
    except ValueError:
        # header doesn't contain names we expect -> fall back to defaults
        idx_survived, idx_pclass, idx_sex, idx_age = 1, 2, 4, 5
else:
    idx_survived, idx_pclass, idx_sex, idx_age = 1, 2, 4, 5

for row in reader:
    if not row:
        continue
    # guard for short rows
    if len(row) <= max(idx_survived, idx_pclass, idx_sex, idx_age):
        continue

    survived = safe_strip(row[idx_survived])
    pclass = safe_strip(row[idx_pclass])
    sex = safe_strip(row[idx_sex]).lower()
    age = safe_strip(row[idx_age])
```

```python
        # We're interested in deceased (Survived == '0' or numeric 0)
        if survived in ('0', '0.0', 'No', 'no', ''):
            # male average age: supply age and count=1 only if age is present and numeric
            if sex == 'male':
                try:
                    a = float(age)
                    # Key name for male average aggregator
                    print(f"male_deceased\t{a}\t1")
                except:
                    # skip males with missing/non-numeric age
                    pass

            # female deceased counts by class (count only; class should be 1/2/3)
            if sex == 'female':
                # normalize class to integer-like (safe)
                pc = pclass
                if pc == '':
                    pc = 'unknown'
                print(f"female_deceased_class_{pc}\t1")
```

REDUCER :

```python
#!/usr/bin/env python3
# reducer.py
import sys

current_key = None

# For male average
male_total_age = 0.0
male_total_count = 0

# For female per-class counts: dict keyed by class string
female_counts = {}

def emit_results():
    # Emit average male age (if any)
    if male_total_count > 0:
        avg = male_total_age / male_total_count
        print(f"Average_age_deceased_males\t{avg:.6f}\t(count={male_total_count})")
    else:
        print("Average_age_deceased_males\tNA\t(count=0)")

    # Emit female deceased counts per class in ascending class order when possible
    if female_counts:
        for pc in sorted(female_counts.keys(), key=lambda x: (x!='unknown', x)):  # put
'unknown' last
            print(f"Female_deceased_class_{pc}\t{female_counts[pc]}")
    else:
```

```python
        print("No_female_deceased_records_found")

for line in sys.stdin:
    line = line.strip()
    if not line:
        continue
    parts = line.split('\t')
    key = parts[0]

    if key == 'male_deceased':
        # male lines: male_deceased\t<age>\t1
        if len(parts) >= 3:
            try:
                age = float(parts[1])
                cnt = int(parts[2])
            except:
                continue
            male_total_age += age
            male_total_count += cnt
    elif key.startswith('female_deceased_class_'):
        # female lines: female_deceased_class_<class>\t1
        # When streamed/sorted, multiple occurrences will come separately
        try:
            cnt = int(parts[1]) if len(parts) > 1 else 1
        except:
            cnt = 1
        pc = key.replace('female_deceased_class_', '')
        female_counts[pc] = female_counts.get(pc, 0) + cnt
    else:
        # ignore other keys (if any)
        pass

# after reading all input, output final aggregates
emit_results()
```

OUTPUT :

File information - part-00000 ✕

Download   Head the file (first 32K)   Tail the file (last 32K)

Block information --   Block 0 ⌄

Block ID: 1073741834

Block Pool ID: BP-965296662-172.16.9.238-1759897177712

Generation Stamp: 1010

Size: 128

Availability:

- ailcomp20

File contents

```
Average_age_deceased_males    30.351111    (count=225)
Female_deceased_class_1   2
Female_deceased_class_2   4
Female_deceased_class_3   37
```

Close