Name: Yashdeep Singh Vilkhu
Student ID: 224195064

1 Function: get_time_series

1.1 Code Implementation

```python
import pandas as pd

def get_time_series(file_path):
    try:
        df = pd.read_csv(file_path)
    except Exception as e:
        raise ValueError(f"Error reading dataset: {e}")

    if "created_at" in df.columns:
        df["created_at"] = pd.to_datetime(df["created_at"], errors="coerce", utc=True)
        df = df.dropna(subset=["created_at"])
        df = df.set_index("created_at").sort_index()

        if "entry_id" in df.columns:
            df["entry_id"] = pd.to_numeric(df["entry_id"], errors="coerce").astype("Int64")

        for col in [c for c in df.columns if c.startswith("field")]:
            df[col] = pd.to_numeric(df[col], errors="coerce")

    elif "time" in df.columns:
        df = df.dropna(subset=["time"])
        df = df.set_index("time").sort_index()

        for col in [c for c in df.columns if c.startswith("s")]:
            df[col] = pd.to_numeric(df[col], errors="coerce")

    else:
        raise ValueError("Dataset must contain either 'created_at' or 'time' column")

    df = df.ffill().bfill()

    return df
```

Listing 1: Python implementation of get-time-series()

1.2 Explanation

- The function reads a CSV file safely with exception handling.
- It supports two dataset formats:
  - `created_at` + `entry_id` + `field1...fieldN` (timestamp-based).

  - `time` + `s1, s2, s3...` (numeric time index).

- If `created_at` exists:
  - Converts it to datetime format.
  - Drops invalid or missing timestamps.
  - Sets `created_at` as index and sorts chronologically.
  - Ensures `entry_id` is integer.
  - Converts `field` columns to numeric.
- If `time` exists:
  - Drops missing values in `time`.
  - Sets `time` as index and sorts.
  - Converts `s` columns to numeric.
- If neither `created_at` nor `time` is found → raises an error.
- Finally, it handles missing values using forward fill (`ffill`) and backward fill (`bfill`) to ensure continuity.
- Returns a clean, time-indexed DataFrame ready for time series analysis.