In [147…
```
import pandas as pd
import datetime
import random
```

In [8]:
```
df = pd.read_csv('../data/processed/processed_oil_prod_data.csv')

df.head()
```

Out[8]:

	field	start_date	name	manifold	sous_manifold	prod	heur	maprod	
0	NORD	2019-08-02	WID-OML121	MF001	SMF001	0.00	0.00	60.54	
1	NORD	2019-08-02	WID-OMLZ233	MF002	SMF002	0.00	0.00	8.53	
2	NORD	2019-08-02	WID-OMLZ173	MF003	SMF003	18.76	14.42	16.68	LAV_F
3	NORD	2019-08-02	WID-OMLZ872	MF002	SMF002	0.00	0.00	19.35	PROD CHA
4	NORD	2019-08-02	WID-OML86	MF004	SMF004	188.38	24.00	0.00	

Gas-Lift types

In [16]:
```
df['gl'].unique()
```

Out[16]:
```
array(['Eruptif', 'Gaz Lift Continu', 'Ex_Gaz Lift', 'Jet pump',
       'ESP pump', 'Gaz Lift Interm.'], dtype=object)
```

We will map each code used to refer to a gas-lift type in the dataset to its corresponding English label, as well as a brief explanation of said type.

In [72]:
```
mappings = {
  "Eruptif": "Natural flow, meaning the well is flowing without any artificial l
  'Gaz Lift Continu': "Continuous gas lift, meaning Gas is continuously injected
  "Ex_Gaz Lift": "A previously Gas Lifted Well, meaning it used to have gas lift
  "Jet pump": "A hydraulic jet pump well, meaning it uses high-pressure fluid in
  "ESP pump": "An Electric Submersible Pump well, meaning it uses an electric pu
  "Gaz Lift Interm.": "An intermittent gas lift, meaning gas is injected intermi
}
```

In [31]:
```
df['start_date']
```

Out[31]:
```
0          2019-08-02
1          2019-08-02
2          2019-08-02
3          2019-08-02
4          2019-08-02
              ...
1848926    2024-03-05
1848927    2024-03-05
1848928    2024-03-05
1848929    2024-03-05
1848930    2024-03-06
Name: start_date, Length: 1848931, dtype: object
```

Start date and well activity per day

We will encode these two features into text format for better interpretability by the LLM

In [96]:
```
def format_date(date):
  date_ = pd.to_datetime(date)
  return datetime.datetime.strftime(date_, "%B %d, %Y")
```

In [111…
```
format_date(df['start_date'].iloc[0])
```

Out[111…
```
'August 02, 2019'
```

In [165…
```
def format_time_period(period):
  if period == 24:
    return "24 hours"
  if period == 0:
    return "0 hours"
  time = str(datetime.timedelta(hours=period)).split(':')[:2]
  return f"{time[0]} hours and {time[1]} minutes"
```

In [116…
```
format_time_period(df['heur'].iloc[2])
```

Out[116…
```
'14 hours and 25 minutes'
```

Text template

In [247…
```
def create_row_template(row):
  date = format_date(row['start_date'])
  field = 'Northern' if row['field'] == 'NORD' else 'Southern'
  gl = mappings[row['gl']]
  time = format_time_period(row['heur']) if row['heur'] > 0 else None
  injection_rate = str(row['injection_rate']) if row['injection_rate'] > 0 else


  return f"""**Well maintenance report assistant**\n\n**Role**: You are a knowle
```

Applying to sample rows

In [248…
```
templates = df.sample(5).apply(create_row_template, axis=1)
```

```
for template in templates:
  print(template, end="\n\n++++++++++++++++++++++++++++++++++++++++++++++++\n\n")
```

Well maintenance report assistant

Role: You are a knowledgible oil and gas maintenance expert at a gas company. You are responsible for the maintenance of gas wells in the oil field. Your task is to analyze well maintenance data and answer technical queries accurately.

Maintenance report
The maintenance activity started on February 17, 2022. The name of the well is WID-ONM152. It belongs to the Northern section of the oil field. The well belongs to manifold MF015 and submanifold SMF046.
On the day of the maintenance activity, The well was not working, and it was producing an average of 0.0 barrels per day. The type of gas lift for this well is Continuous gas lift, meaning Gas is continuously injected into the well to assist in lifting the fluids. Gas has been injected into the well at a rate of 78783.46m3/day. The value of the well's valve is 18.09. The pressure value at the top and bottom of the well are 32.91 and 30.42 respectively. The well is located in the zone with code 11 of the field. And the associated reservoir rock has a permeability coefficient of 8.5175

++

Well maintenance report assistant

Role: You are a knowledgible oil and gas maintenance expert at a gas company. You are responsible for the maintenance of gas wells in the oil field. Your task is to analyze well maintenance data and answer technical queries accurately.

Maintenance report
The maintenance activity started on August 20, 2020. The name of the well is WID-OMP18. It belongs to the Northern section of the oil field. The well belongs to manifold MF001 and submanifold SMF001.
On the day of the maintenance activity, The well was working for 24 hours, and it was producing an average of 39.42 barrels per day. The type of gas lift for this well is Natural flow, meaning the well is flowing without any artificial lift. No gas has been injected into the well. The value of the well's valve is 19.96. The pressure value at the top and bottom of the well are 26.08 and 10.51 respectively. The well is located in the zone with code 8 of the field. And the associated reservoir rock has a permeability coefficient of 1.5381424

++

Well maintenance report assistant

Role: You are a knowledgible oil and gas maintenance expert at a gas company. You are responsible for the maintenance of gas wells in the oil field. Your task is to analyze well maintenance data and answer technical queries accurately.

Maintenance report

The maintenance activity started on February 27, 2021. The name of the well is WI
D-OMNZ833. It belongs to the Northern section of the oil field. The well belongs
to manifold MF004 and submanifold SMF064.
On the day of the maintenance activity, The well was working for 24 hours, and it
was producing an average of 49.19 barrels per day. The type of gas lift for this
well is Natural flow, meaning the well is flowing without any artificial lift. No
gas has been injected into the well. The value of the well's valve is 29.68. The
pressure value at the top and bottom of the well are 82.26 and 45.57 respectivel
y. The well is located in the zone with code HZN of the field. And the associated
reservoir rock has a permeability coefficient of 8.649373

++

Well maintenance report assistant

Role: You are a knowledgible oil and gas maintenance expert at a gas company.
You are responsible for the maintenance of gas wells in the oil field. Your task
is to analyze well maintenance data and answer technical queries accurately.

Maintenance report
The maintenance activity started on July 19, 2021. The name of the well is WID-HG
A32. It belongs to the Southern section of the oil field. The well belongs to man
ifold MF031 and submanifold SMF100.
On the day of the maintenance activity, The well was not working, and it was prod
ucing an average of 0.0 barrels per day. The type of gas lift for this well is Na
tural flow, meaning the well is flowing without any artificial lift. No gas has b
een injected into the well. The value of the well's valve is 10.36. The pressure
value at the top and bottom of the well are 30.78 and 25.09 respectively. The wel
l is located in the zone with code HZP of the field. And the associated reservoir
rock has a permeability coefficient of 0.71589

++

Well maintenance report assistant

Role: You are a knowledgible oil and gas maintenance expert at a gas company.
You are responsible for the maintenance of gas wells in the oil field. Your task
is to analyze well maintenance data and answer technical queries accurately.

Maintenance report
The maintenance activity started on January 07, 2020. The name of the well is WID
-MDZ661. It belongs to the Southern section of the oil field. The well belongs to
manifold MF025 and submanifold SMF109.
On the day of the maintenance activity, The well was working for 24 hours, and it
was producing an average of 16.8 barrels per day. The type of gas lift for this w
ell is Natural flow, meaning the well is flowing without any artificial lift. No
gas has been injected into the well. The value of the well's valve is 19.08. The
pressure value at the top and bottom of the well are 74.42 and 53.22 respectivel
y. The well is located in the zone with code 13 of the field. And the associated
reservoir rock has a permeability coefficient of 8.55575

++

In-Context Learning: Incorporating examples into prompts

In [220…

```
def example_templates(row):

  examples = [
    {
      "question": f"What was the cause of maintenance of well {row['name']} on {
      "answer": f"The cause of maintance of well {row['name']} on {format_date(r
    },
    {
      "question": f"How many barrels of oil was the well producing on {format_da
      "answer": f"On {format_date(row['start_date'])}, the well was producing {r
    },
    {
      "question": f"How long was the well working on {format_date(row['start_dat
      "answer": f"The well was working for {format_time_period(row['heur'])} on
    },
    {
      "question": f"What is the type of gas lift for well {row['name']}?",
      "answer": f"The type of gas lift for well {row['name']} is {mappings[row['
    },
    {
      "question": f"What is the value of the well {row['name']}'s valve on {form
      "answer": f"The value of the well's valve is {row['duze']}"
    },
    {
      "question": f"What is the pressure at the top of the well {row['name']} on
      "answer": f"The pressure at the top of the well is equal to {row['pr_tet']
    },
    {
      "question": f"What is the pressure at the bottom of the well {row['name']}
      "answer": f"The pressure at the bottom of the well is equal to {row['press
    },
    {
      "question": f"What is the permeability coefficient of the reservoir rock a
      "answer": f"The permeability coefficient of the reservoir rock associated
    }
  ]

  return examples
```

In [163…

```
def generate_examples(df, n_examples=3):
  examples = []
  samples = df.sample(n=n_examples)
  for sample in samples.iterrows():
    examples.append(random.choice(example_templates(sample[1])))

  return examples
```

In [171…
```
generate_examples(df, 5)
```

Out[171…
```
[{'question': 'What was the cause of maintenance of well WID-MD94 on February 0
4, 2020?',
  'answer': 'The cause of maintance of well WID-MD94 on February 04, 2020 is No
rmal'},
 {'question': 'What is the pressure at the top of the well WID-MDZ555 on Octobe
r 31, 2019',
  'answer': 'The pressure at the top of the well is equal to 19.69'},
 {'question': "What is the value of the well WID-OMKZ212's valve on February 0
1, 2020?",
  'answer': "The value of the well's valve is 36.29"},
 {'question': 'What is the pressure at the top of the well WID-OMOZ801 on March
05, 2022',
  'answer': 'The pressure at the top of the well is equal to 31.93'},
 {'question': 'What was the cause of maintenance of well WID-OMPZ34 on August 0
8, 2020?',
  'answer': 'The cause of maintance of well WID-OMPZ34 on August 08, 2020 is No
rmal'}]
```

In [221…
```
generate_examples(df, 1)
```

Out[221…
```
[{'question': 'What is the pressure at the top of the well WID-MD338 on Decembe
r 31, 2019?',
  'answer': 'The pressure at the top of the well is equal to 0.0'}]
```

Generating the prompt

In [239…
```
def generate_prompt(row):
  row_template = create_row_template(row)
  examples = generate_examples(df)
  examples_text = "\n".join([f"Question: {example['question']}\nAnswer: {example
  prompt = row_template + f"""**Examples:**\n {examples_text}\n\n
  **Task:**\n\nYour task is to answer the following question: \n"""

  return prompt
```

In [240…
```
print(generate_prompt(df.iloc[99]))
```

```
**Well maintenance report assistant**

**Role**: You are a knowledgible oil and gas maintenance expert at a gas company.
You are responsible for the maintenance of gas wells in the oil field. Your task
is to analyze well maintenance data and answer technical queries accurately.

---

**Maintenance report**

  The maintenance activity started on August 02, 2019. The name of the well is WI
D-OMM643. It belongs to the Northern section of the oil field. The well belongs t
o manifold MF005 and submanifold SMF008.
On the day of the maintenance activity, The well was not working, and it was prod
ucing an average of 0.0 barrels per day. The type of gas lift for this well is Na
tural flow, meaning the well is flowing without any artificial lift. No gas has b
een injected into the well. The value of the well's valve is 8.11. The pressure v
alue at the top and bottom of the well are 28.64 and 23.18 respectively. The well
is located in the zone with code 1C of the field. And the associated reservoir ro
ck has a permeability coefficient of 1.04

---

  **Examples:**
 Question: What is the pressure at the top of the well WID-MDZ559 on March 08, 20
20?
Answer: The pressure at the top of the well is equal to 60.19

Question: What was the cause of maintenance of well WID-OMN451 on May 06, 2021?
Answer: The cause of maintance of well WID-OMN451 on May 06, 2021 is DIV_FERME GO
R ELEVE

Question: What is the pressure at the top of the well WID-MDZ585 on August 17, 20
20?
Answer: The pressure at the top of the well is equal to 27.8



  **Task:**

Your task is to answer the following question:
```

Generating QA pairs for training

In [232…

```python
def generate_question_answer_pairs(row):
  qa_pair = random.choice(example_templates(row))

  input = generate_prompt(row) + f"\n{qa_pair['question']}"
  output = qa_pair['answer']

  return {
    "input": input,
    "output": output
  }
```

In [269…
```
sample = df.sample().apply(generate_question_answer_pairs, axis=1).iloc[0]

sample
```

Out[269…
```
{'input': "**Well maintenance report assistant**\n\n**Role**: You are a knowled
gible oil and gas maintenance expert at a gas company. You are responsible for
the maintenance of gas wells in the oil field. Your task is to analyze well mai
ntenance data and answer technical queries accurately.\n\n---\n\n**Maintenance
report**\nThe maintenance activity started on September 10, 2019. The name of t
he well is WID-ONI551. It belongs to the Northern section of the oil field. The
well belongs to manifold MF014 and submanifold SMF042.\nOn the day of the maint
enance activity, The well was working for 24 hours, and it was producing an ave
rage of 45.86 barrels per day. The type of gas lift for this well is Continuous
gas lift, meaning Gas is continuously injected into the well to assist in lifti
ng the fluids. Gas has been injected into the well at a rate of 32007.77m3/day.
The value of the well's valve is 24.86. The pressure value at the top and botto
m of the well are 26.98 and 22.78 respectively. The well is located in the zone
with code HZN of the field. And the associated reservoir rock has a permeabilit
y coefficient of 1.3905932\n\n---\n**Examples:**\n Question: How long was the w
ell working on September 30, 2022?\nAnswer: The well was working for 24 hours o
n September 30, 2022\n\nQuestion: What is the pressure at the top of the well W
ID-MDZ717 on March 13, 2021?\nAnswer: The pressure at the top of the well is eq
ual to 27.44\n\nQuestion: What is the permeability coefficient of the reservoir
rock associated with well WID-OMP75?\nAnswer: The permeability coefficient of t
he reservoir rock associated with well WID-OMP75 is 1.084539\n\n\n\n  **Task:**
\n\nYour task is to answer the following question: \n\nWhat is the permeability
coefficient of the reservoir rock associated with well WID-ONI551?",
 'output': 'The permeability coefficient of the reservoir rock associated with
well WID-ONI551 is 1.3905932'}
```