```
import pandas as pd

# Load CSV data
csv_path = "customer_survey.csv"
survey_df = pd.read_csv(csv_path)


# Preview data
print("Original CSV Data:")
print(survey_df.head())
```

```
Original CSV Data:
   customer_id         name                    email region  rating  \
0            1     John Doe     john.doe@example.com  North       5
1            2   Jane Smith   jane.smith@example.com   West       4
2            3  Emily Davis  emily.davis@example.com  South       3

            comments review_date
0    Great product!  2024-01-01
1    Fast delivery.  2024-01-02
2  Average quality.  2024-01-03
```

```
# Drop rows with missing values in key fields
survey_df.dropna(subset=['customer_id', 'rating', 'review_date'], inplace=True)

# Convert date column to datetime
survey_df['review_date'] = pd.to_datetime(survey_df['review_date'], errors='coerce')

# Filter out invalid ratings
survey_df = survey_df[survey_df['rating'].between(1, 5)]


import os

# Create directory if it doesn't exist
os.makedirs('/mnt/data', exist_ok=True)

# Then save the file
clean_csv_path = "/mnt/data/cleaned_customer_survey.csv"
survey_df.to_csv(clean_csv_path, index=False)

print("\nCleaned CSV saved for bulk insert.")
```

```
Cleaned CSV saved for bulk insert.
```

```
import json

# Load JSON data
json_path = "web_feedback.json"
with open(json_path, 'r') as file:
    web_feedback = json.load(file)

# Convert to DataFrame
web_df = pd.DataFrame(web_feedback)

# Clean data
web_df = web_df.dropna(subset=['customer_id', 'rating'])
web_df = web_df[web_df['rating'].between(1, 5)]

# Save for SQL import
clean_json_path = "/mnt/data/cleaned_web_feedback.csv"
web_df.to_csv(clean_json_path, index=False)
print("Cleaned JSON data saved as CSV for import.")
```

```
Cleaned JSON data saved as CSV for import.
```

```
import xml.etree.ElementTree as ET

# Load and parse XML
xml_path = "external_reviews.xml"
tree = ET.parse(xml_path)
root = tree.getroot()

# Extract data
xml_reviews = []
for review in root.findall('review'):
    try:
        review_data = {
            'customer_id': int(review.find('customer_id').text),
            'rating': int(review.find('rating').text),
            'comments': review.find('comments').text
        }
        if 1 <= review_data['rating'] <= 5:
            xml_reviews.append(review_data)
    except Exception as e:
        print(f"Error parsing review: {e}")

# Convert to DataFrame
xml_df = pd.DataFrame(xml_reviews)

# Save to CSV
clean_xml_path = "/mnt/data/cleaned_external_reviews.csv"
xml_df.to_csv(clean_xml_path, index=False)
print("Cleaned XML data saved as CSV.")
```

```
print("Cleaned XML data saved as CSV.")
```

```
Cleaned XML data saved as CSV.
```

```
from google.colab import files
files.download('/mnt/data/cleaned_customer_survey.csv')
files.download('/mnt/data/cleaned_web_feedback.csv')
files.download('/mnt/data/cleaned_external_reviews.csv')
```