Intel Data Center

— Data Set **Descriptions**

`intel.energy_data`: Contains information about daily energy production and consumption for different regions in the United States.

- `balancing_authority` - A Balancing Authority is responsible for maintaining the electricity balance within its region. This is a company that makes sure electricity is being exchanged between electric providers and regions so that no region runs out of electricity due to high demand.
- `date` - The date the energy was produced.
- `region` - The electric service area within a geographic area of the USA. e.g. California, Midwest, etc.
- `time_at_end_of_hour` - The time and date after energy was generated, .e.g., energy generated between 1pm-2pm will show up as 2pm in this field.
- `demand` - The energy demand in megawatts (MW) on the grid (what the houses/business are using).
- `net_generation` - The energy produced in MW in the region by all sources e.g., wind, coal, nuclear, etc.
- `all_petroleum_products` - The energy produced in MW by petroleum products.
- `coal` - The energy produced in MW by all coal products
- `hydropower_and_pumped_storage` - The energy produced in MW by water power and pumped heat sources.
- `natural_gas` - The energy produced in MW by natural gas sources
- `nuclear` - The energy produced in MW from nuclear fuel sources
- `solar` - The energy produced in MW by solar panels and other solar energy capturing methods.

- `wind` - The energy produced in MW from wind turbines and other wind sources.

`intel.power_plants`: Contains general information about power plants in the United States.

- `plant_name` - The name of the power plant.
- `plant_code` - The unique identifier of the plant.
- `region` - The region in the US where the power plant is located. Matches the regions in the intel.energy_data
- `state` - The state where the power plant is located.
- `primary_technology` - The primary technology used to generate electricity at the power plant.

`intel.energy_by_plant`: Contains total energy production information at the plant for the year 2022.

- `plant_name` - The name of the power plant.
- `plant_code` - The unique identifier of the plant.
- `energy_type` - The kind of energy generated by the power plant. Either renewable energy or fossil fuel.
- `energy_generated_mw` - The total energy generated, in MegaWatts, at the plant for the year 2022.

— **Task 1:** Energy Generation

First things first, I want to write a query using the intel.energy_data table that calculates the sum total of energy produced, and group it by each region. I then want to sort the output by highest total energy. This will allow me to see which region has the highest positive total energy?

```
SELECT
  region,
  SUM(net_generation - demand) AS total_energy
FROM
  intel.energy_data
```

```
GROUP BY
  region
ORDER BY
  total_energy DESC
```

region	total_energy
Mid-Atlantic	31693087
Northwest	21685235
Southwest	14986086
Central	10842350
Southeast	5446206
Texas	-1578632
Carolinas	-4033212
Tennessee	-5959294
Florida	-9591378
Midwest	-15521417
New England	-16718662
New York	-28310973
California	-64939167

It's clear that the **Mid-Atlantic** region produces the highest total energy at 31,693,087 MegaWatts.

The next step in my SQL code is to write a query that calculates the total sum of renewable energy generated by each region, focusing on hydropower_and_pumped_storage, wind, and solar sources. I will then sort the results to identify the top two regions with the highest renewable energy production.

```
SELECT
  region,
  SUM(hydropower_and_pumped_storage + wind + solar) AS
renewable_energy
FROM
  intel.energy_data
GROUP BY
  region
ORDER BY
  renewable_energy DESC
```

region	renewable_energy
Northwest	199266574
Texas	131367234
Central	120536677
Midwest	118303446
California	78255584
Mid-Atlantic	55401624
New York	30952145
Southwest	19471766
Tennessee	19144807
Southeast	19090580
Carolinas	12679733
New England	11882645
Florida	10923319

Northwest and **Texas** are the top two regions for total renewable energy production, at 199,266,574 and 131,367,234 MegaWatts, respectively.

The next step is to modify my query to calculate the percentage of renewable energy generated by each region. This will allow me to see if there are any changes in the ranking of the regions based on their contribution to total renewable energy.

```
SELECT
  region,
  SUM(hydropower_and_pumped_storage + wind + solar) /
SUM(net_generation) * 100 AS pct_renewable_energy
FROM
  intel.energy_data
GROUP BY
  region
ORDER BY
  pct_renewable_energy DESC
```

region	pct_renewable_energy
Northwest	52.25167583605338045
Central	41.01264146599812153
California	37.45315023009070984
Texas	30.61962525194856692
New York	24.88706998143432633
Midwest	16.88281012276880415
Southwest	15.57428162182849462
Tennessee	11.95938117184001791
New England	11.85825647964600644
Southeast	7.748970846976154434
Mid-Atlantic	6.580127274554451620
Carolinas	5.719941319023146257
Florida	4.435845642716237617

Now the top two are **Northwest** and **Central**. In fact, Texas has moved to fourth. More than 52% (52.25%) of energy produced in the Northwest region comes from renewable sources.

– **Task 2:** Generating New Data by Energy Type

Intel is interested in understanding how renewable energy and fossil fuels trend over time. To achieve this I need to write a query that calculates the renewable energy generated for each row, returning only the date, region, and energy_generated_mw columns.

```
SELECT
  date,
  region,
```

```
  hydropower_and_pumped_storage + wind + solar AS
energy_generated_mw,
  'renewable energy' AS energy_type
FROM
  intel.energy_data
```

The next step is to write a query that calculates the fossil fuel energy generated for each row.

```
SELECT
  date,
  region,
  all_petroleum_products + coal + natural_gas + nuclear +
other_fuel_sources AS energy_generated_mw,
  'fossil fuel' AS energy_type
FROM
  intel.energy_data
```

Now I want to write a query that unions both of these together.

```
SELECT
  date,
  region,
  hydropower_and_pumped_storage + wind + solar AS
energy_generated_mw,
  'renewable energy' AS energy_type
FROM
  intel.energy_data
UNION
SELECT
  date,
  region,
```

```
  all_petroleum_products + coal + natural_gas + nuclear +
other_fuel_sources AS energy_generated_mw,
  'fossil fuel' AS energy_type
FROM
  intel.energy_data
```

Task 3: Aggregating Power Plant Data

The next step is to join the `intel.power_plants` and `intel.energy_by_power_plant` tables on the `plant_code`. This joined table will serve as the foundation for further aggregations and analysis, helping the Intel team make informed decisions about the location of their next data center.

```
SELECT
  *
FROM
  intel.energy_by_plant AS ebp
  INNER JOIN intel.power_plants AS pp ON ebp.plant_code =
pp.plant_code
```

To determine the region with the most renewable power plants, I'll write a query that counts the total number of renewable energy power plants in each region. The query will group the results by region and return the region with the highest count. This will help identify which region has the most renewable power plants.

```
WITH joined_data as (
  SELECT
    *
  FROM
    intel.energy_by_plant AS ebp
    INNER JOIN intel.power_plants AS pp ON ebp.plant_code =
pp.plant_code
)
```

```
SELECT
  region,
  count(*) as n_plants
FROM
  joined_data
WHERE
  energy_type = 'renewable_energy'
GROUP BY
  region
ORDER BY
  n_plants DESC
```

region	n_plants
Midwest	234
Northwest	185
Texas	184
California	148
Central	108
Florida	86
Mid-Atlantic	83
New England	70
Carolinas	64
Southwest	58
Southeast	44
New York	18
Hawaii	17

The region with the most renewable energy plants is the **Midwest** region with **234** plants.

The next step is to write a query that returns both the total number of power plants and the total energy generated from plants using "Solar Photovoltaic" technology, grouped by region. This will provide insight into the distribution and output of solar energy across different regions.

```
WITH joined_data as (
  SELECT
    *
  FROM
    intel.energy_by_plant AS ebp
    INNER JOIN intel.power_plants AS pp ON ebp.plant_code =
pp.plant_code
)
SELECT
  region,
  count(*) as n_plants,
  sum(energy_generated_mw) as total_energy_mw
FROM
  joined_data
WHERE
  primary_technology = 'Solar Photovoltaic'
GROUP BY
  region
HAVING
  count(*) >= 50
ORDER BY
  total_energy_mw DESC
```

region	n_plants	total_energy_mw
Texas	57	17425264
California	59	16247303
Northwest	74	11719631
Florida	79	10264795
Carolinas	54	6338560
Midwest	71	4907305

Despite the midwest having the 3rd largest total number of plants that use solar technology,they are producing about 40% of the total energy that Texas generates. This is either likely due to having older technology at the plants or having smaller power plants that use this technology.

– **Task 4:** Hourly Trends in Renewable Energy

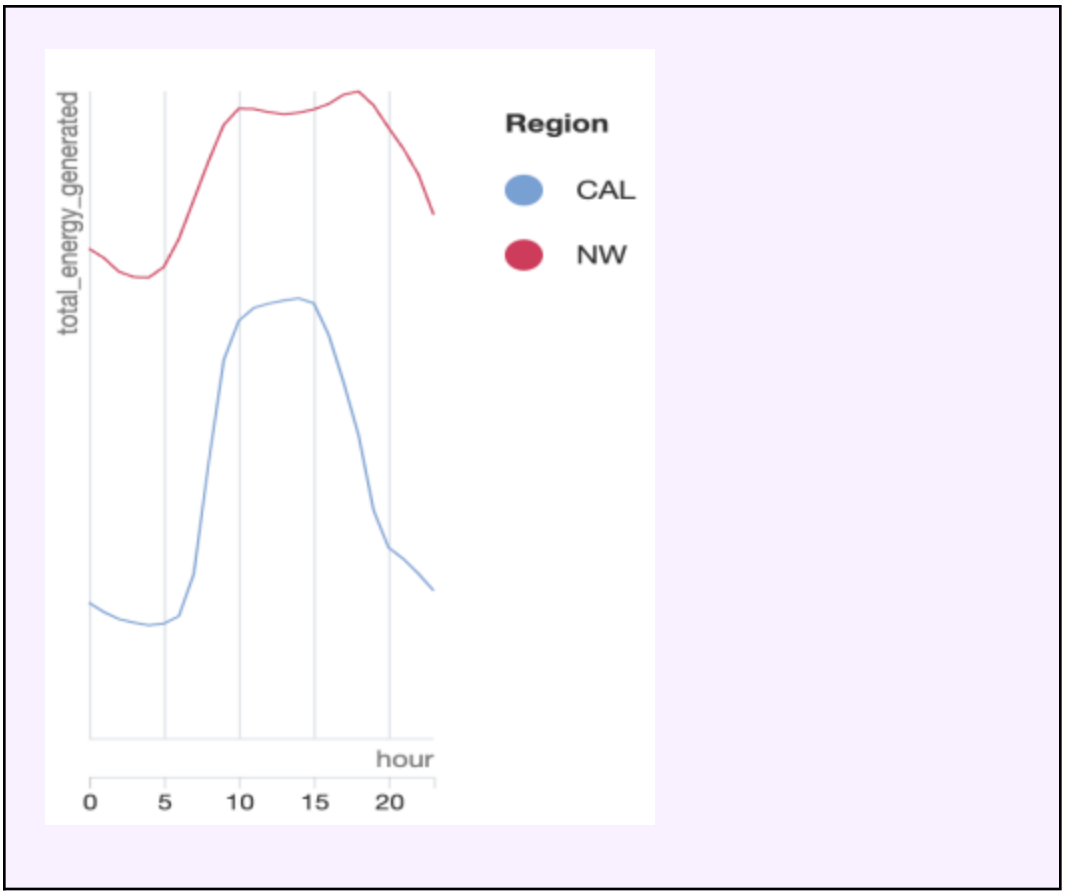
Before diving into the Tableau visualizations, the next step is to write a query that calculates the total renewable energy generated in each region for each hour of the day. This will allow us to investigate how renewable energy generation fluctuates throughout the day across different regions.

```
SELECT
  date_part('hour', time_at_end_of_hour) as hour,
  region,
  SUM(hydropower_and_pumped_storage + wind + solar) AS
total_energy_generated
FROM
  intel.energy_data
GROUP BY
  region,
  hour
```

Looking at the data, I want to now modify the query to filter to the ‘California’ and ‘Northwest’ regions only.

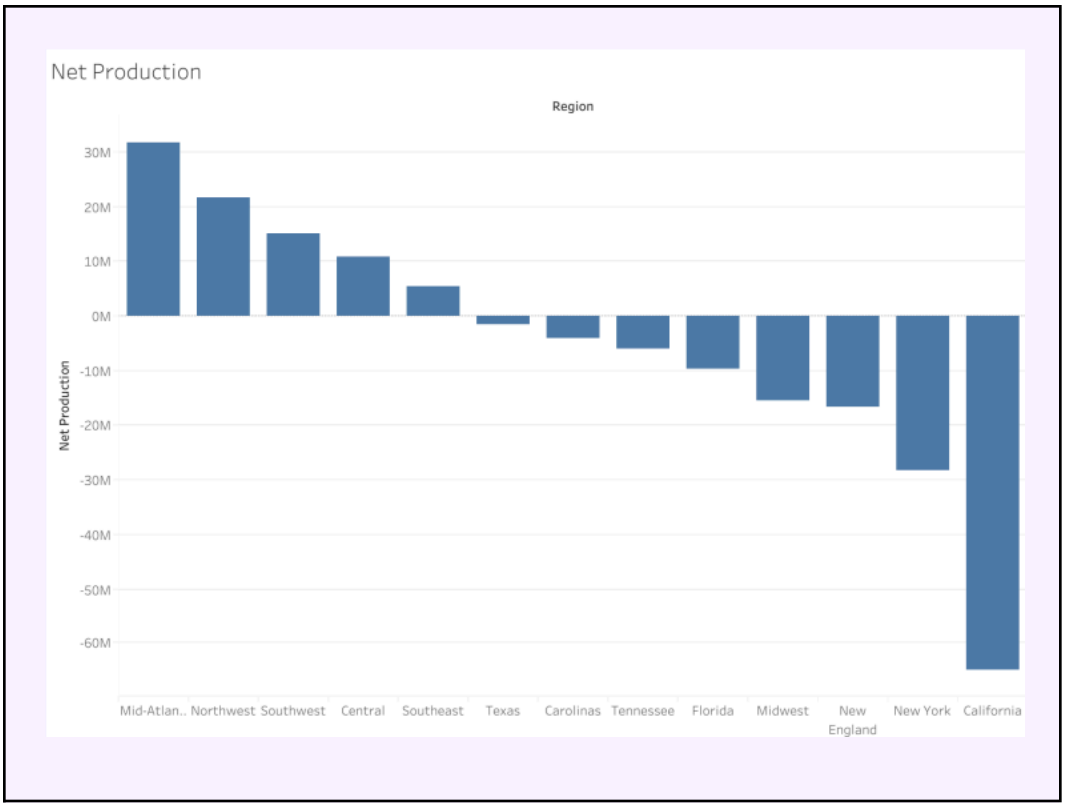
```
SELECT
  date_part('hour', time_at_end_of_hour) as hour,
  region,
  SUM(hydropower_and_pumped_storage + wind + solar) AS
total_energy_generated
FROM
  intel.energy_data
WHERE
  region in ('California', 'Northwest')
GROUP BY
  region,
  hour
ORDER BY
  region,
  hour
```

Using the SQL built in visualizer I decided to plot a line graph of the energy generated for each hour of the day and color it by region.

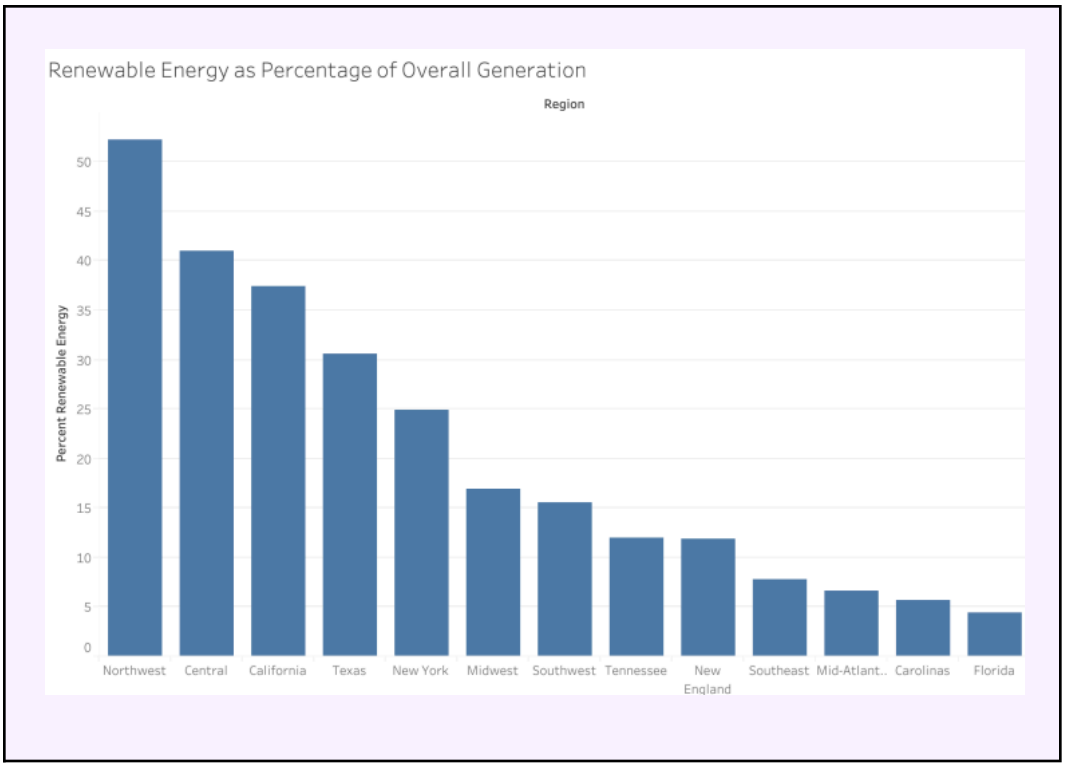


– **Task 5:** Visualizing and Analyzing Using Tableau

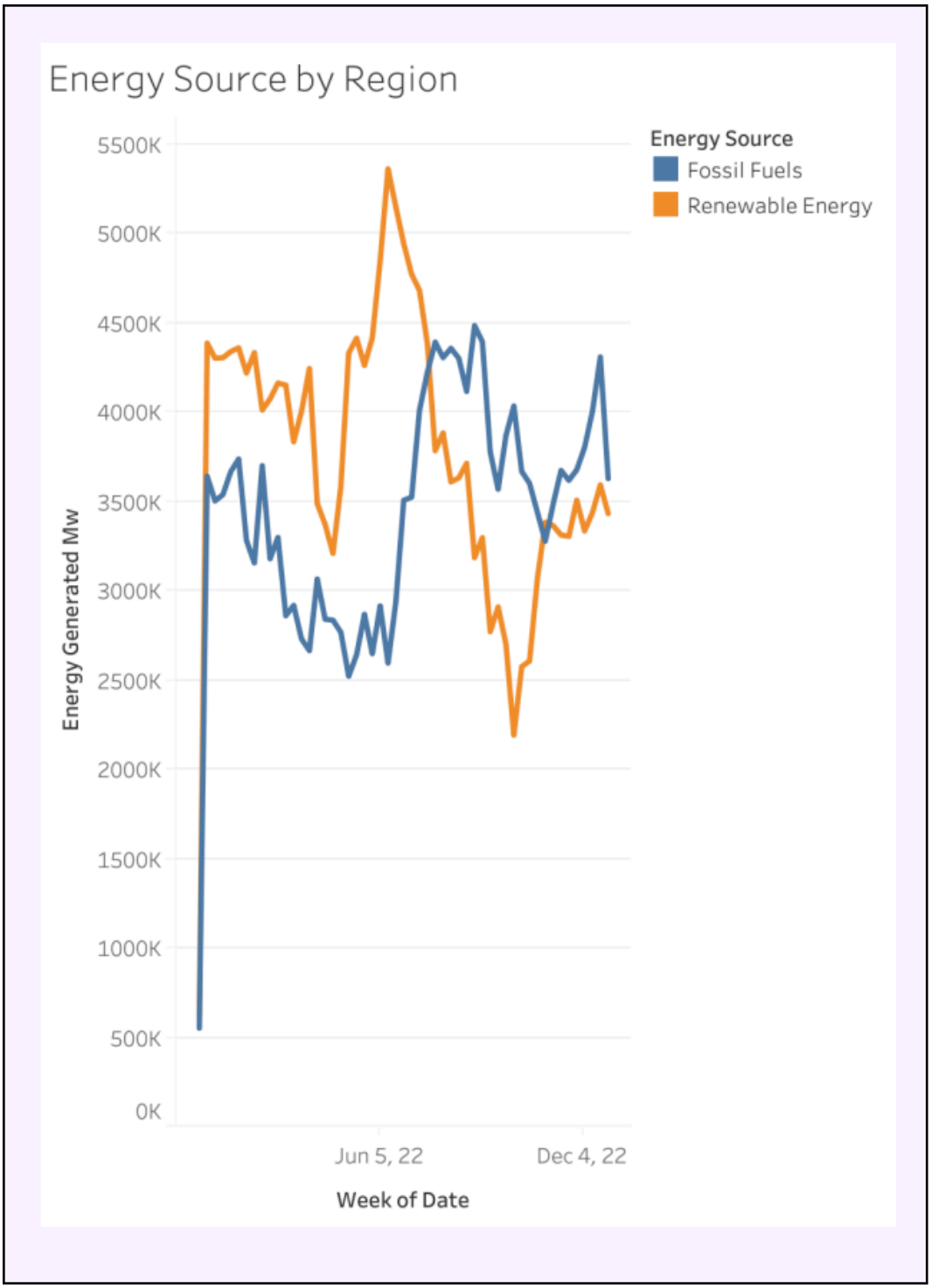
Now, moving on to Tableau, the next step is to create a bar chart. This chart will display the net production by region, sorted in descending order from tallest to smallest. The visualization will help identify which regions have the highest net energy production, guiding Intel's decision on the best location for their next data center.



The next step is to create a bar chart that shows the percentage of renewable energy generated by each region. This will help identify the regions that are leading in renewable energy production, which is crucial for Intel's decision on where to place their next data center.

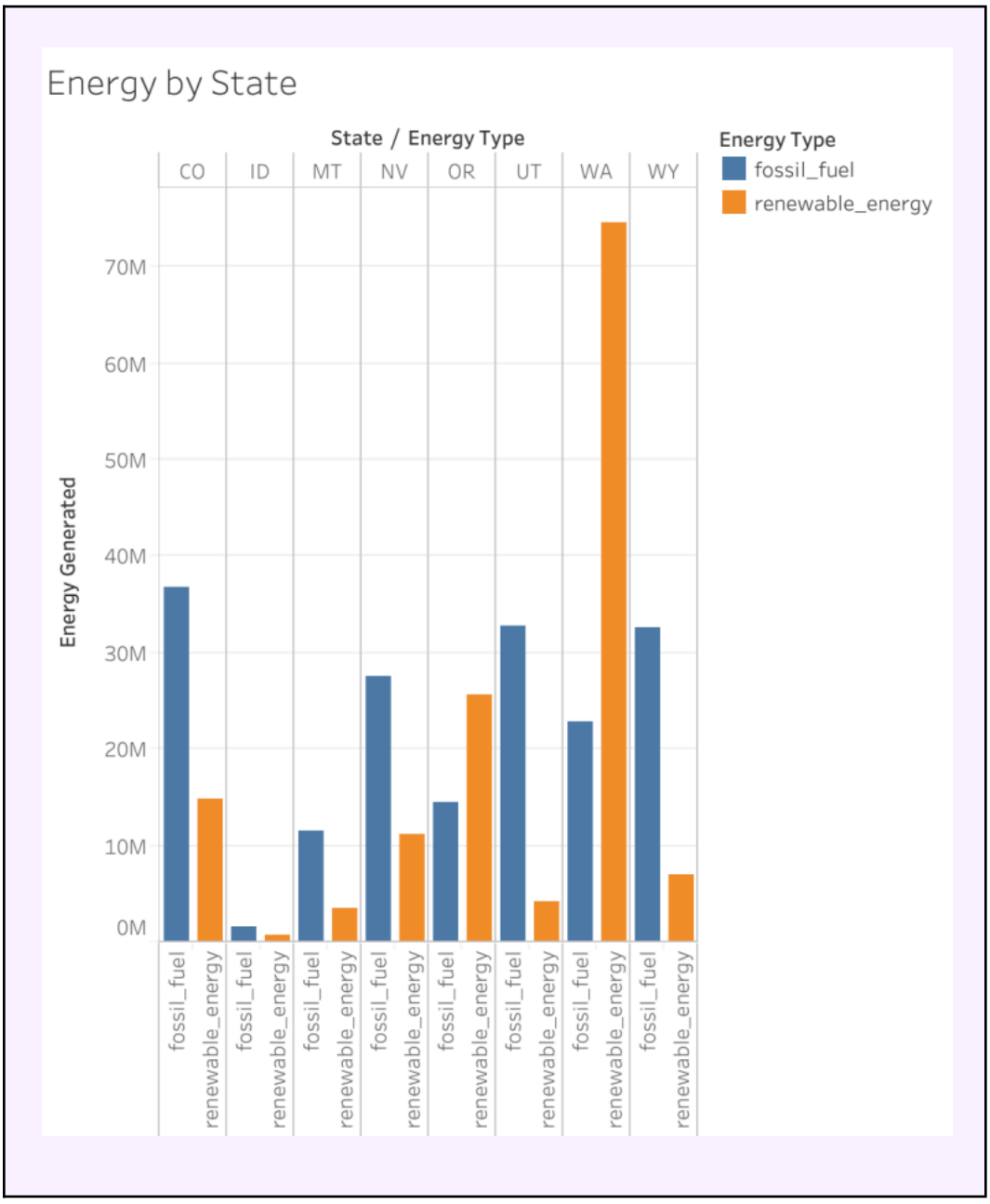


The next step is to create a line chart showing the energy generated for each energy source—fossil fuels and renewable energy—at the weekly date level. I'll also add a filter for the region to make the chart more interactive and allow for region-specific analysis.



The next step is to create a bar chart displaying the total energy generated by each state, broken down by energy type. I'll color the bars by energy type for clear differentiation and include a region filter to reduce the number of bars shown,

making the chart easier to analyze region-specific data.



– **Task 5:** Communicating Results

When looking at all of the data I believe that the Northwest region would be the ideal location to build a new data center. As for the state, I believe that we need more information to make a more informed decision, but based on what I have seen, I think that Washington would be a good fit. The Northwest region is a significant net energy producer with a surplus of 21 million megawatts; this is ideal for Intel because this showcases that this region is committed to sustainable energy sources. The Northwest also has the highest percentage of renewable energy at 52.2% which only further promotes the idea that this region has a strong emphasis on clean and sustainable power sources. It also has a relatively consistent supply of energy which is a must when determining the compatibility. When looking at the state of Washington, it has a high percentage of renewable energy as well. The state is known for its abundant hydropower resources which is a reliable and consistent source of energy. All in all, with a substantial surplus of energy, a focus on renewable sources, and a robust infrastructure, the Northwest region and Washington in particular provides scalability and support for future growth.