Project | Sustainability Impact Analysis for Intel

INTRODUCTION: As you learned listening in on the strategy meeting with Dr. Alvarez and Intel's Sustainability Team, Intel is committed to reducing its carbon footprint and improving the sustainability of its devices – not just during manufacturing, but throughout the entire lifecycle.

A key part of this effort is their repurposing programs, which play a central role in achieving these sustainability goals. Repurposing and recycling programs aim to reduce e-waste, energy consumption, and CO_2 emissions by extending the life of existing devices, and thus reducing the need for new device manufacturing. Like Michael Campbell said: the average household in the US has anywhere from 3–5 PCs devices, tablets, notebooks, desktops that are perfectly functional, but not being used!

One challenge Intel faces is determining which devices in its repurposing program should be prioritized for the maximum environmental benefit. That's where data analysis comes in! To help with this, Intel gathered data on each device repurposed or recycled in 2024.

Your task is to evaluate the effectiveness of Intel's current repurposing strategy and provide a data-driven recommendation to help guide the program's direction and optimize sustainability efforts.

HOW IT WORKS: Follow the prompts in the questions below to investigate the data. Post your answers in the provided boxes: the **yellow boxes** for the queries you write and **blue boxes** for your text-based analysis. Once you're done, you'll submit your ***completed*** .pdf file to HQ for feedback from The Accelerator Team.

SQL App: Here's the link to our specialized SQL app, where you'll write your SQL queries and interact with the data.

NOTE: The dataset you are working with is designed for The Global Career Accelerator to reflect the key characteristics and structure of Intel's real data, while protecting their confidentiality and proprietary information. Be aware that any conclusions or results derived from this dataset should be viewed as hypothetical and for illustrative purposes only.

— Data Set **Descriptions**

In this project you'll query 2 different datasets, `intel.device_data` and `intel.impact_data`, that you will join together for your analysis. Here you'll find the data dictionary for each dataset.

`intel.device_data`

- `device_id`: Unique identifier for each repurposed device
- `device_type`: Type of device, values are either "Laptop" or "Desktop"
- `model_year`: The year the device was manufactured (e.g., 2018, 2019, etc.)

`intel.impact_data`

- `impact_id`: Unique identifier for the repurposed device's impact record (e.g., "LP20NA141592")
- `device_id`: Unique identifier linking the impact record to a specific device in the `intel.device_data` table
- `usage_purpose`: The specific purpose for which the device is being repurposed, values are Education & Digital Literacy, Corporate & Enterprise, Government & Public Sector, Environmental Sustainability Programs, and Social Impact & Non-Profit
- `power_consumption`: Power consumption of the device in watts (W) when in use (e.g., 50W, 75W)
- `energy_savings_yr`: Estimated energy savings per device per year when repurposed compared to a new device, measured in kilowatt-hours (kWh)
- `co2_saved_kg_yr`: Estimated CO2 emissions saved per device per year from manufacturing a new device, measured in kilograms (kg).
- `recycling_rate`: The percentage of the device that is recyclable (e.g., 80%, 90%).
- `region`: The geographical region where the device was repurposed, values are "North America", "Europe", and "Asia"

— **Task 1:** Organizing and Understanding the Data

We'll start by **joining** the device data with the impact data, allowing for a comprehensive analysis of device types, model years, repurpose regions, and energy savings in one dataset.

A. Simply write a query that returns all of the columns from both tables, joining the two on the `device_id` column. Be sure to choose the appropriate join so that all relevant data is included in your result.

(paste your query below 👇)

```
select
 *
from
 intel.device_data as d
 full outer join intel.impact_data as i
 on d.device_id = i.device_id
```

B. To your joined dataset, add a new column called `device_age` calculated by subtracting the model_year from 2024. Paste your query below and double check that the values in your new column make sense. For example, a 2019 device should be 5 years old.

(paste your query below 👇)

```
with joined as (select
 *,
 2024 -d.model_year as device_age
from
 intel.device_data as d
 full outer join intel.impact_data as i
 on d.device_id = i.device_id)
```

C. Order your joined data by model_year (oldest to newest). Do you notice more older (5+ years) or newer (under 5 years) devices being repurposed? What might that indicate?

(write your **answer** below 👇)

I see more newer devices, which could maybe be indicative of a larger supply.

D. Bucketing the `device_age` will allow us to analyze trends and patterns in energy savings and CO_2 reductions more effectively than using individual ages. Use a `CASE WHEN` clause to add one more column, called `device_age_bucket`, to your data, that is based on the device_age:

- `WHEN` the `device_age` is less than or equal to 3, `device_age_bucket` should be "newer"
- `WHEN` the `device_age` is greater than 3 but less than or equal to 6, `device_age_bucket` should be "mid-age"
- `WHEN` the `device_age` is greater than 6, `device_age_bucket` should be "older"

HINT: Instead of using e.g. `device_age <= 3`, you need to reference the calculation directly: `2024 - d.model_year <= 3`.

Double check that the values in your new column make sense! For example, a 2019 device should be characterized as "mid-age".

(paste your query below 👇)

```
with joined as (
  select
    *,
    2024 - d.model_year as device_age,
    case
      when 2024 - d.model_year <= 3 then 'newer'
      when 2024 - d.model_year > 3 and 2024 -
d.model_year <= 6 then 'mid-age'
      when 2024 - d.model_year > 6 then 'older'
    end as device_age_bucket
  from intel.device_data as d
  full outer join intel.impact_data as i
```

```
    on d.device_id = i.device_id
)

select
 *
from
 joined
```

```
with joined as (
  select
    *,
    2024 - d.model_year as device_age,
    case
      when 2024 - d.model_year <= 3 then 'newer'
      when 2024 - d.model_year > 3 and 2024 - d.model_year <= 6 then 'mid-age'
      when 2024 - d.model_year > 6 then 'older'
```

150000 rows | Incomplete | 5.718 seconds

act_id	usage_purpose	power_consumption	energy_savings_yr	co2_saved_kg_yr	recycling_rate	region	device_age	device_age_bucket
9AS7449301044	Education & Digital Literacy	78	31.0	18.6	0.94	Asia	5	mid-age
1EU5846101045	Government & Public Sector	81	15.1	3.8	0.94	Europe	3	newer
2NA7472001045	Corporate & Enterprise	75	16.5	6.6	0.95	North America	2	newer
1NA3412001048	Government & Public Sector	50	21.9	8.8	0.88	North America	3	newer
1EU4471901048	Corporate & Enterprise	82	29.1	7.3	0.95	Europe	3	newer
6AS2512401051	Education & Digital Literacy	60	35.1	21.1	0.85	Asia	8	older
7AS6840701052	Education & Digital Literacy	61	49.2	29.5	0.87	Asia	7	older
0NA6864301053	Corporate & Enterprise	52	47.1	18.8	0.86	North America	4	mid-age
0NA2669901056	Education & Digital Literacy	82	45.4	18.2	0.97	North America	4	mid-age
2AS2961601057	Government & Public Sector	79	5.7	3.4	0.94	Asia	2	newer
0AS7306101063	Corporate & Enterprise	77	37.4	22.4	0.94	Asia	4	mid-age
1AS7650501063	Education & Digital Literacy	81	6.3	3.8	0.93	Asia	3	newer
9NA5114401064	Corporate & Enterprise	81	34.0	13.6	0.97	North America	5	mid-age

— **Task 2:** Key Insights

Now it's time to analyze the overall impact of Intel's repurposing program. You will use your final query from **Task 1** together with the `WITH` keyword for the remainder of this Project as you aggregate and analyze the data you've organized and prepped. For a refresher, rewatch "🍿The `WITH` Keyword" in SkillBuilder 6.

A. What is the total number of devices Intel repurposed in 2024?

HINT: The dataset **is** representing all devices repurposed in 2024! You just need to COUNT all the rows in your joined data from Task 1!

(write your **answer** below 👇)

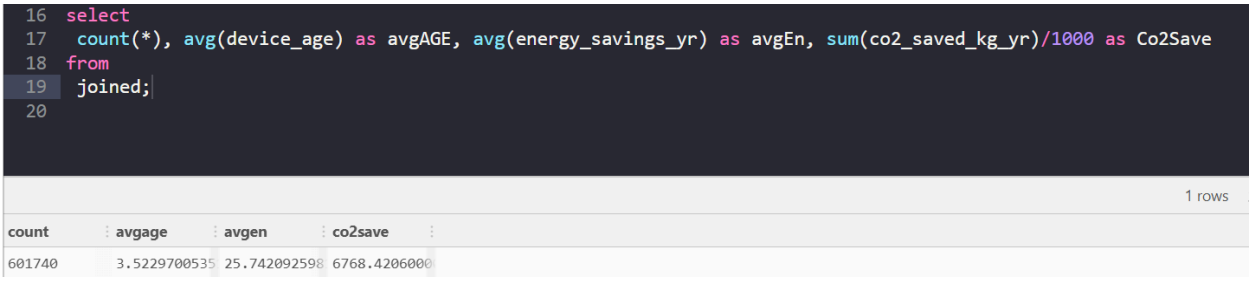
There are 601,740 devices being repurposed in 2024.

B. Write a query that returns the total number of devices repurposed, the average age of repurposed devices in 2024, the average estimated energy savings (kWh) from repurposed devices per year, and the total CO_2 emissions saved (in tons) from repurposed devices.

Note: CO_2 emissions are typically measured in tons. Since `CO₂_saved_kg_yr` is measured in kg, divide the `SUM(CO₂_saved_kg_yr)` by 1000 to report the total CO_2 emissions saved in tons.

(paste your query below 👇)

```
select
 count(*), avg(device_age) as avgAGE,
avg(energy_savings_yr) as avgEn,
sum(co2_saved_kg_yr)/1000 as Co2Save
from
 joined
```



```
16  select
17   count(*), avg(device_age) as avgAGE, avg(energy_savings_yr) as avgEn, sum(co2_saved_kg_yr)/1000 as Co2Save
18  from
19   joined;
20
```

1 rows

count	avgage	avgen	co2save
601740	3.5229700535	25.742092598	6768.4206000

C. Now that you have calculated the average estimated energy savings (kWh) and CO_2 emissions saved (tons), use ChatGPT to help put these numbers into perspective.

Try this prompt: I found that each repurposed device saves approximately of XXX kWh of energy per year and Intel's repurposing program saved XXX tons of CO_2 emissions in one year. Help me understand the significance of these

numbers. How would this compare to the energy consumption of a small city or the amount of CO_2 produced by cars? What is the environmental impact of these savings?

What comparisons did you find most impactful in terms of scale? Summarize how much energy and CO_2 emissions were saved and how it compares to something familiar, like powering households or reducing car emissions.

(write your **answer** below 👇)

25.74 kWh of energy were saved per a device that is the same as an LED bulb burning for 2 and a half years straight. 6,768.42 metric tons of C02 were saved which is comparable to burning 7.5 million pounds of coal. Each refurbished device having an average age of 3.5 years has a significant impact.

— **Task 3:** Identifying Trends & Maximizing Sustainability

By grouping our data in different ways, we can uncover patterns in energy savings and CO_2 reductions. These insights will help us determine which categories of devices contribute the most to sustainability efforts and where Intel should focus its repurposing strategy for maximum impact.

A. Write a query that returns the total number of devices, the average energy savings, and the average CO_2 emissions saved (in tons), grouped by device_type.

Note (again): You'll need to divide `AVG(CO₂_saved_kg_yr)` by 1000 to report the average CO_2 emissions saved in tons.

(paste your query below 👇)

```
select
  device_type,
  avg(energy_savings_yr) as avgEn,
  avg(co2_saved_kg_yr) / 1000 as Co2avg,
  count(*)
```

```
from
  joined
group by
  device_type
```

B. Based on the results, which device type contributes the most to energy savings and CO_2 reduction? Why might that be the case?

Hint: Don't forget you can use ChatGPT as your Teammate to help think through your response!

(write your **answer** below 👇)

Laptops tend to contribute more to energy savings and c02 reduction. Laptops use low-voltage cpus and compact storage designed to optimally complete tasks not brute forcing them like desktops.

C. Write a query that returns the total number of devices, the average energy savings, and the average CO_2 emissions saved (in tons), now grouped by `device_age_bucket`.

(paste your query below 👇)

```
select
 device_age_bucket,
 avg(energy_savings_yr) as avgEn,
 avg(co2_saved_kg_yr) / 1000 as Co2avg,
 count(*)

from
 joined
group by
 device_age_bucket
```

D. Based on the result of your query, what do you notice about the relationship between device age and the number of devices repurposed versus the average energy saved?

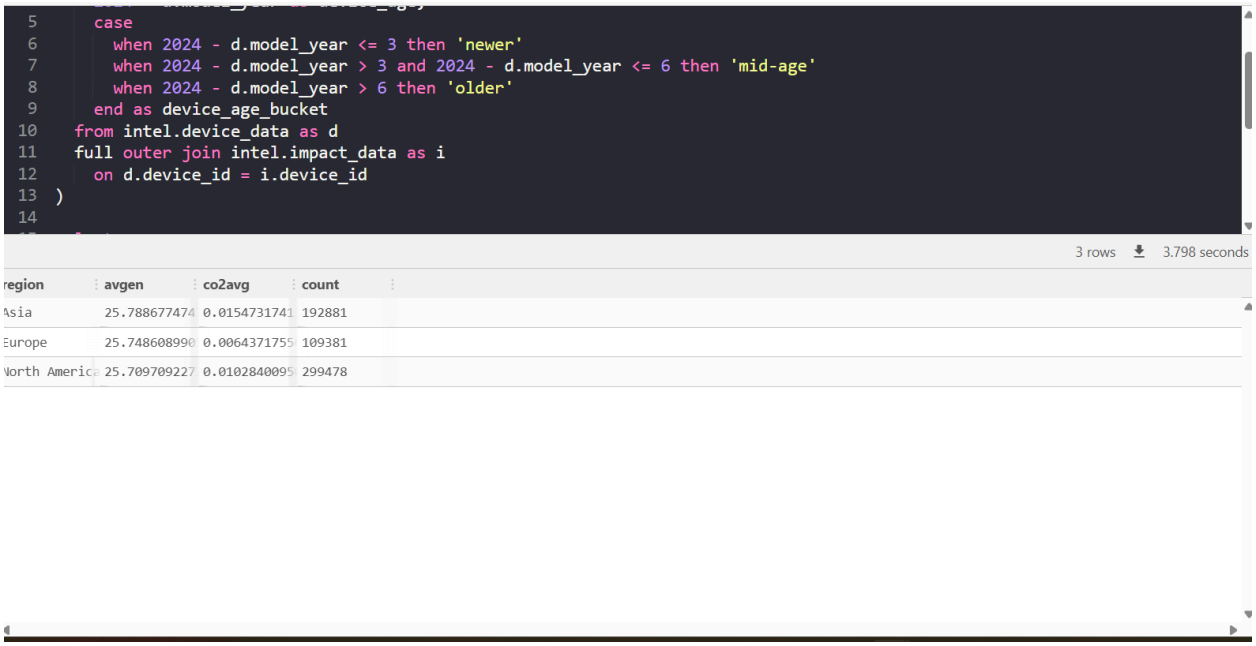
(write your **answer** below 👇)

There is a positive correlation between age and energy savings. While there is a negative correlation between age and number of devices.

E. Finally, write a query that returns the total number of devices, the average energy savings, and the average CO_2 emissions saved (in tons), now grouped by region.

(paste your query below 👇)

```
--3rd sc for github
select
  region,
  avg(energy_savings_yr) as avgEn,
  avg(co2_saved_kg_yr) / 1000 as Co2avg,
  count(*)
from
  joined
group by
  region
```



```
  case
    when 2024 - d.model_year <= 3 then 'newer'
    when 2024 - d.model_year > 3 and 2024 - d.model_year <= 6 then 'mid-age'
    when 2024 - d.model_year > 6 then 'older'
  end as device_age_bucket
from intel.device_data as d
full outer join intel.impact_data as i
  on d.device_id = i.device_id
)
```

3 rows 3.798 seconds

region	avgen	co2avg	count
Asia	25.788677474	0.0154731741	192881
Europe	25.74860899	0.0064371755	109381
North Americ	25.709709227	0.0102840095	299478

F. How does the carbon intensity of electricity in each region impact the total CO_2 savings from repurposed devices? Are there regions where repurposing leads to significantly higher environmental benefits? Why might that be?

(write your **answer** below 👇)

> Yes Europe has similar energy saving stats compared to the other two and has a lower C02 avg savings. However it has approximately 100k less devices then Asia and 200k less devices than North America. Europe's energy grid is cleaner, which is why the C02avg savings are lower.

— **Task 4:** Data-Driven Recommendations

Using the findings from this analysis, we need to summarize key takeaways and develop actionable recommendations for Intel. Remember: the goal is to refine Intel's repurposing strategy to maximize energy savings and CO_2 reductions while ensuring the most effective use of resources.

A. Based on your analysis of the repurposed devices (including energy savings, CO_2 emissions, and device age), write **four** key takeaways in succinct sentences/bullets that summarize the most important patterns and insights from the data. These should be specific, concise, and focused on the implications of repurposing newer versus older devices.

(write your **answer** below 👇)

- More than half the devices in the intel dataset are in the 'newer' age bucket category, which could be indicative of a larger supply.
- Only 3.36% of computers are older computers that have been refurbished.
- It is more efficient to refurbish mid -age devices compared to newer ones by a factor of 20.09%. (energy to C02 savings)/ number of devices.
- Older devices save 13x more c02 per a device compared to newer ones.

B. Based on your four key takeaways and ChatGPT as your teammate, write a recommendation for Intel on how to improve the repurposing program. Your recommendation should include a clear action or strategy for Intel based on the data and a data-driven justification for why this approach would maximize energy savings and CO_2 reductions.

(write your **answer** below 👇)

I think that Intel should focus more on trying to refurbish more midlife and old age laptops. The data shows that the older the computer the more of an impact it will have on the environment. This is because before refurbishing, older computers have less efficient hardware which chews through power quickly. I still think they should continue refurbishing new-age computers, because there is a large supply. However, they should try to get the majority of devices from Asia or North America where it will have more of an impact.

C. Briefly reflect on how ChatGPT's suggestions influenced your recommendation. Did it help you see something you hadn't considered? What parts of your recommendation were improved based on its response?

(write your **answer** below 👇)

One thing ChatGpt suggested was a real world application of my strategy. It suggested establishing legacy hardware recovery partnerships. It said we should try and target schools, government agencies, and enterprise clients to try and recover old computers in exchange for Intel store credit.

– **LevelUp:** Optimizing Repurposing Strategy for Maximum Impact

Now that you've gained insights into the energy savings and CO_2 reductions across different device types and regions, let's use this data to optimize Intel's repurposing strategy for maximum environmental benefit.

A. Add to your final query of Task 3 that returns the total number of devices, the average energy savings, and the average CO_2 emissions saved (in tons), grouped by region, **the percentage** of the total energy savings and CO_2 reductions contributed by each device type within each region.

HINT: To calculate the percentage of the total energy savings, use this formula:
`Total energy savings for the device type / Total energy savings for the region) * 100`
You'll use a similar one for the percentage of the total CO_2 reductions.

Try this prompt: What's the best way to calculate the percentage of CO_2 reductions contributed by each device type in each region?

(paste your query below 👇)

```
Paste your query here.
```

B. Based on the results of your query, analyze the data to answer:

- Which device types in which regions contribute the most energy savings and CO_2 reductions relative to their numbers?

- How can this analysis help Intel prioritize specific device types in certain regions to maximize environmental benefits?

(write your **answer** below 👇)

Write your answer here.

C. In addition to focusing on sustainability, imagine Intel needs to optimize for cost-effectiveness in their repurposing program. How might you adjust your query to incorporate cost data (e.g., cost per repurposed device)? What strategies could Intel use to balance sustainability goals with cost constraints?

(write your **answer** below 👇)

Write your answer here.