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. When done properly, the result should have 601,740 rows.

(paste your query below 👇)

```
SELECT
  *
FROM
  intel.device_data as a FULL
  OUTER JOIN intel.impact_data as b on a.device_id =
b.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 👇)

```
SELECT
  *,
  2024 - model_year as device_age
FROM
  intel.device_data as a FULL
  OUTER JOIN intel.impact_data as b on a.device_id =
b.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 👇)

The majority of devices in this dataset are newer, with the majority of repurposed models being from 2020 to 2022. This indicates that people are upgrading their models within 4 years.

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 👇)

```
SELECT
  *,
  CASE
    WHEN 2024 - model_year <= 3 THEN 'newer'
    WHEN 2024 - model_year <= 6 THEN 'mid-age'
    ELSE 'older'
  END AS device_age_bucket
FROM
  intel.device_data as a FULL
  OUTER JOIN intel.impact_data as b on a.device_id =
b.device_id
```

— **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?

(write your **answer** below 👇)

> Intel repurposed 601,740 devices 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 👇)

```
WITH device_age AS(
  SELECT
    *,
    2024 - model_year as device_age,
    CASE
      WHEN 2024 - model_year <= 3 THEN 'newer'
      WHEN 2024 - model_year <= 6 THEN 'mid-age'
      ELSE 'older'
    END AS device_age_bucket
  FROM
    intel.device_data as a FULL
    OUTER JOIN intel.impact_data as b on a.device_id =
b.device_id
```

```
)
SELECT
  COUNT(*) as total_devices,
  AVG(device_age) as avg_age,
  AVG(energy_savings_yr) as avg_energy_savings,
  SUM(co2_saved_kg_yr) / 1000 as co2_saved
FROM
  device_age
```

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 👇)

The most impactful comparison Chat made was the CO2 saved being the equivalent of 1,472 gasoline cars off the road for a year. This makes the data seem much more significant than a number with no context.

— **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 👇)

```
WITH device_age AS(
  SELECT
    *,
    2024 - model_year as device_age,
    CASE
      WHEN 2024 - model_year <= 3 THEN 'newer'
      WHEN 2024 - model_year <= 6 THEN 'mid-age'
      ELSE 'older'
    END AS device_age_bucket
  FROM
    intel.device_data as a FULL
    OUTER JOIN intel.impact_data as b on a.device_id =
b.device_id
)
SELECT
  device_type,
  COUNT(*) as total_devices,
  AVG(energy_savings_yr) as avg_energy_savings,
  AVG(co2_saved_kg_yr) / 1000 as avg_co2_saved
FROM
  device_age
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 contribute the most to total energy savings and CO2 saved solely because there are twice as many laptops being repurposed as desktops. In terms of averages, energy savings and CO2 saved are nearly identical.

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 👇)

```
WITH device_age AS(
  SELECT
    *,
    2024 - model_year as device_age,
    CASE
      WHEN 2024 - model_year <= 3 THEN 'newer'
      WHEN 2024 - model_year <= 6 THEN 'mid-age'
      ELSE 'older'
    END AS device_age_bucket
  FROM
    intel.device_data as a FULL
    OUTER JOIN intel.impact_data as b on a.device_id =
b.device_id
)
SELECT
  device_age_bucket,
  COUNT(*) as total_devices,
  AVG(energy_savings_yr) as avg_energy_savings,
  AVG(co2_saved_kg_yr) / 1000 as avg_co2_saved
FROM
  device_age
```

```
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 👇)

> This query demonstrates that repurposing older devices saves significantly more energy and CO2 than newer 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 👇)

```
WITH device_age AS(
  SELECT
    *,
    2024 - model_year as device_age,
    CASE
      WHEN 2024 - model_year <= 3 THEN 'newer'
      WHEN 2024 - model_year <= 6 THEN 'mid-age'
      ELSE 'older'
    END AS device_age_bucket
  FROM
    intel.device_data as a FULL
    OUTER JOIN intel.impact_data as b on a.device_id =
b.device_id
)
SELECT
  region,
  COUNT(*) as total_devices,
  AVG(energy_savings_yr) as avg_energy_savings,
```

```
  AVG(co2_saved_kg_yr) / 1000 as avg_co2_saved
FROM
  device_age
GROUP BY
  region
```

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 👇)

CO2 saved is highest in Asia, then North America, and then Europe, where average energy savings are essentially the same. This is because the energy used in Asia and North America is largely carbon-based, so repurposing makes a larger difference than Europe, where energy used is from renewable sources.

— **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 👇)

Repurposing older devices is better than repurposing newer devices.
Repurposing devices in Asia is more beneficial than repurposing devices in Europe.

There is no significant difference between laptops and desktops when it comes to the benefits of repurposing.
Repurposing is beneficial under any circumstances.

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 👇)

The benefits of repurposing devices can be maximized by focusing on older devices in countries like Asia and North America where carbon-based energy is more frequently used.

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 👇)

Without using ChatGPT I would not have been able to come up with why Asia and North America benefit from saving CO2 emissions more than Europe. With that knowledge, my argument was stronger and I feel more confident in my findings.

— **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.