Climate-Based Sales Analysis: Fans and Cooling Devices

Temperature Impact on Product Sales in Swiss Cities

1 Executive Summary

This analysis investigates the relationship between average temperature and sales of fans and cooling devices in eight Swiss cities. The central research question is: **Is there a correlation between higher temperatures and increased sales of cooling products?**

Main hypothesis: In warmer regions (especially in Ticino, e.g., Lugano with 24.8°C), more fans and cooling devices are sold than in cooler cities.

2 1. Business Problem & Unit of Analysis

2.1 1.1 Problem Statement

Digitec Galaxus wants to understand whether and how strongly climatic conditions influence demand for cooling products.

2.2 1.2 Unit of Analysis

The analysis unit is **Swiss cities**. We examine the following eight cities:

1. **Zürich** (ZH) - 436,551 residents, 21.1°C
2. **Bern** (BE) - 137,995 residents, 20.7°C
3. **Luzern** (LU) - 86,234 residents, 21.2°C
4. **Basel** (BS) - 177,571 residents, 22.2°C
5. **St. Gallen** (SG) - 78,863 residents, 19.6°C
6. **Lugano** (TI) - 63,629 residents, 24.8°C (warmest city)
7. **Lausanne** (VD) - 144,873 residents, 22.4°C

8. **Genf** (GE) - 209,061 residents, 22.5°C

The average temperature refers to the **warmest month** of the year.

3 2. Data Preparation

3.1 2.1 Load Packages

```
# Load packages
library(tidyverse)
library(readr)
library(ggplot2)
library(knitr)
library(scales)
library(corrplot)

# Set options
options(scipen = 999)
theme_set(theme_minimal())
```

3.2 2.2 Load Data

```
# Load temperature and population data
temp_data <- read_delim("Temparatur und Anzahl Personen.csv",
                        delim = ";",
                        locale = locale(encoding = "UTF-8"))

# Rename columns
colnames(temp_data) <- c("City", "Population", "Temperature")

# Load sales data
sales_data <- read_csv("DigitecLive_Cleaned.csv",
                       locale = locale(encoding = "UTF-8"))

kable(temp_data,
      caption = "Cities with Population and Average Temperature",
      format.args = list(big.mark = "'"))
```

Table 1: Cities with Population and Average Temperature

City	Population	Temperature
Zürich	436'551	21.1
Bern	137'995	20.7
Luzern	86'234	21.2
Basel	177'571	22.2
St. Gallen	78'863	19.6
Lugano	63'629	24.8
Lausanne	144'873	22.4
Genf	209'061	22.5

3.3 2.3 Filter Cooling Products

We filter products with the following keywords in category or product name:

- **Fans**: ventilator, lüfter, fan
- **Air conditioners**: klimaanlage, klimagerät, klima, cool, cooling

```
# Filter cooling products based on category and product name
cooling_keywords <- c(
  "ventilator", "lüfter", "luefter", "fan",
  "klimaanlage", "klimagerät", "klimageraet", "klima",
  "cool", "cooling", "kühlung", "kuehlung"
)

# Create regex pattern
pattern <- paste(cooling_keywords, collapse = "|")

# Filter relevant products
cooling_sales <- sales_data %>%
  filter(
    str_detect(tolower(`infos.Category`), pattern) |
    str_detect(tolower(fullProductName), pattern)
  ) %>%
  # Exclude irrelevant categories
  filter(
    !str_detect(tolower(`infos.Category`), "wasserkuehlung|wasserkühlung|cpu-"),
    !str_detect(tolower(`infos.Category`), "pc-luefter|pc-lüfter"),
    !str_detect(tolower(`infos.Category`), "luftreiniger"),
    !str_detect(tolower(fullProductName), "luftreiniger|purifier"),
```

```
    !str_detect(tolower(`infos.Category`), "game-|gaming-controller-zubehoer|heizluefter|hei
  )

# Filter for target cities
target_cities <- c("Zürich", "Bern", "Luzern", "Basel",
                   "St. Gallen", "Lugano", "Lausanne", "Genf")

cooling_sales_filtered <- cooling_sales %>%
  filter(cityName_clean %in% target_cities)

cat(sprintf("Filtered products in 8 cities: %d (%.2f%%)\n",
            nrow(cooling_sales_filtered),
            100 * nrow(cooling_sales_filtered) / nrow(sales_data)))
```

```
Filtered products in 8 cities: 296 (0.20%)
```

4 3. Data Analysis

4.1 3.1 Sales by City

```
# Aggregate sales by city
sales_by_city <- cooling_sales_filtered %>%
  group_by(cityName_clean) %>%
  summarise(
    Total_Sales = n(),
    Total_Revenue = sum(`salesPrice.amountIncl`, na.rm = TRUE),
    Avg_Price = mean(`salesPrice.amountIncl`, na.rm = TRUE),
    .groups = 'drop'
  ) %>%
  rename(City = cityName_clean)

# Merge with temperature data
analysis_data <- sales_by_city %>%
  left_join(temp_data, by = "City")

# Calculate sales per 5,000 residents
analysis_data <- analysis_data %>%
  mutate(Sales_per_5k = (Total_Sales / Population) * 5000)
```

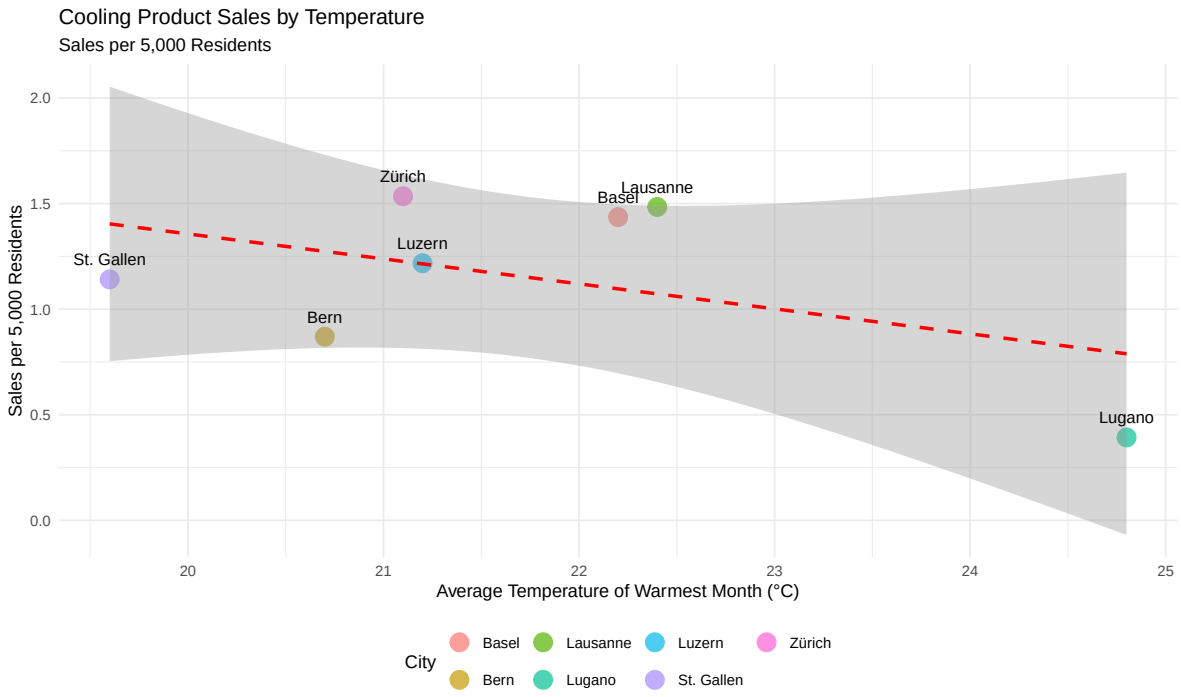
```r
kable(analysis_data %>%
        select(City, Temperature, Population, Sales_per_5k) %>%
        arrange(desc(Sales_per_5k)),
      digits = 2,
      format.args = list(big.mark = "'"),
      caption = "Sales per 5,000 Residents by City")
```

Table 2: Sales per 5,000 Residents by City

City	Temperature	Population	Sales_per_5k
Zürich	21.1	436'551	1.53
Lausanne	22.4	144'873	1.48
Basel	22.2	177'571	1.44
Luzern	21.2	86'234	1.22
St. Gallen	19.6	78'863	1.14
Bern	20.7	137'995	0.87
Lugano	24.8	63'629	0.39

4.2 3.2 Visualization: Sales vs. Temperature

```r
# Normalized sales vs. temperature
ggplot(analysis_data, aes(x = Temperature, y = Sales_per_5k)) +
  geom_point(aes(color = City), size = 5, alpha = 0.7) +
  geom_smooth(method = "lm", se = TRUE, color = "red", linetype = "dashed") +
  geom_text(aes(label = City), vjust = -1.2, size = 3.5) +
  labs(
    title = "Cooling Product Sales by Temperature",
    subtitle = "Sales per 5,000 Residents",
    x = "Average Temperature of Warmest Month (°C)",
    y = "Sales per 5,000 Residents",
    color = "City"
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")
```



5 4. Correlation Analysis

5.1 4.1 Pearson Correlation

```
# Correlation between temperature and normalized sales
cor_result <- cor.test(analysis_data$Temperature,
                       analysis_data$Sales_per_5k,
                       method = "pearson")

cat("CORRELATION ANALYSIS\n")
```

```
CORRELATION ANALYSIS
```

```
cat("====================\n\n")
```

```
====================
```

```r
cat(sprintf("Correlation coefficient (r): %.3f\n", cor_result$estimate))
```

```
Correlation coefficient (r): -0.479
```

```r
cat(sprintf("p-value: %.4f\n", cor_result$p.value))
```

```
p-value: 0.2767
```

```r
cat(sprintf("Interpretation: %s\n",
            ifelse(cor_result$p.value < 0.05,
                   "Significant correlation (p < 0.05)",
                   "No significant correlation (p >= 0.05)")))
```

```
Interpretation: No significant correlation (p >= 0.05)
```

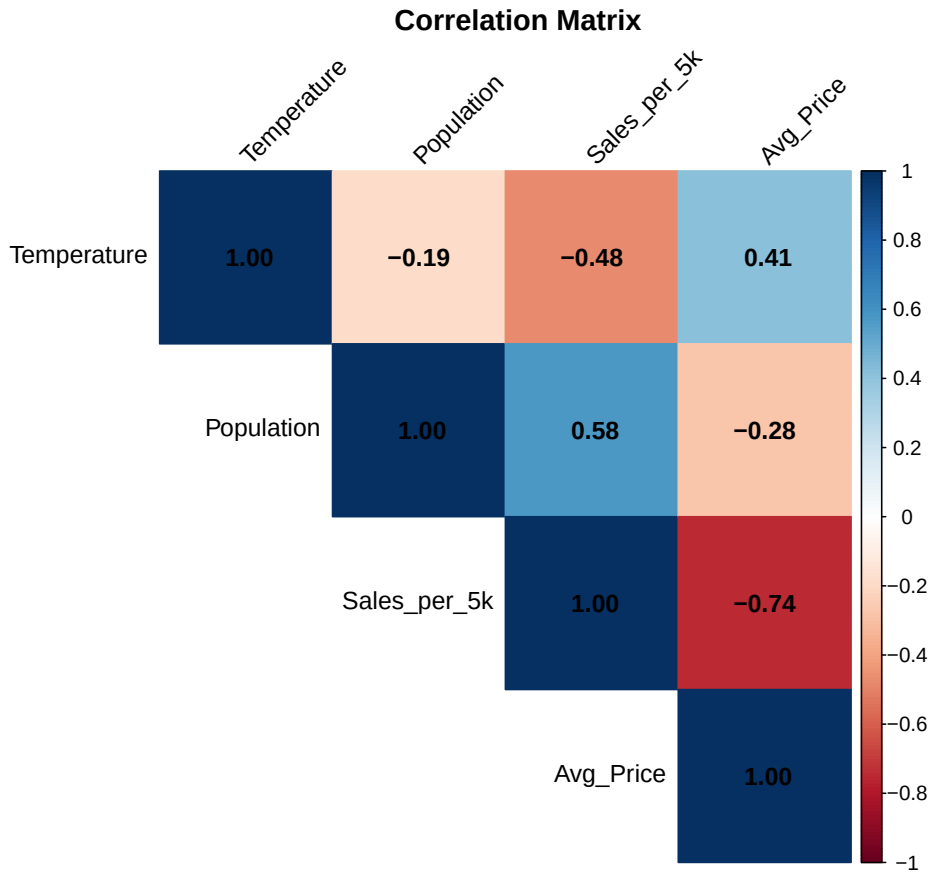
5.2 4.2 Linear Regression

```r
# Linear model: normalized sales ~ temperature
model <- lm(Sales_per_5k ~ Temperature, data = analysis_data)
summary(model)
```

```

Call:
lm(formula = Sales_per_5k ~ Temperature, data = analysis_data)

Residuals:
        1         2         3         4         5         6         7 
 0.339761 -0.404132  0.411433 -0.395813  0.003037 -0.262634  0.308348 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.72245    2.10973   1.764    0.138
Temperature -0.11830    0.09692  -1.221    0.277

Residual standard error: 0.3921 on 5 degrees of freedom
Multiple R-squared:  0.2296,    Adjusted R-squared:  0.07547 
F-statistic:  1.49 on 1 and 5 DF,  p-value: 0.2767
```

5.3 4.3 Correlation Matrix

```
# Correlation matrix
cor_data <- analysis_data %>%
  select(Temperature, Population, Sales_per_5k, Avg_Price)

cor_matrix <- cor(cor_data, use = "complete.obs")

corrplot(cor_matrix,
         method = "color",
         type = "upper",
         addCoef.col = "black",
         tl.col = "black",
         tl.srt = 45,
         title = "Correlation Matrix",
         mar = c(0, 0, 2, 0))
```



6 5. City Comparison

6.1 5.1 City Ranking

```
# Ranking by temperature and sales
ranking <- analysis_data %>%
  mutate(
    Temp_Rank = rank(-Temperature),
    Sales_Rank = rank(-Sales_per_5k)
  ) %>%
  select(City, Temperature, Temp_Rank, Sales_per_5k, Sales_Rank) %>%
  arrange(Temp_Rank)

kable(ranking,
      digits = 2,
      format.args = list(big.mark = "'"),
      col.names = c("City", "Temp (°C)", "Temp Rank",
                    "Sales per 5k", "Sales Rank"),
      caption = "City Ranking by Temperature and Sales")
```

Table 3: City Ranking by Temperature and Sales

City	Temp (°C)	Temp Rank	Sales per 5k	Sales Rank
Lugano	24.8	1	0.39	7
Lausanne	22.4	2	1.48	2
Basel	22.2	3	1.44	3
Luzern	21.2	4	1.22	4
Zürich	21.1	5	1.53	1
Bern	20.7	6	0.87	6
St. Gallen	19.6	7	1.14	5

6.2 5.2 Ticino vs. Other Cities

```
# Lugano (Ticino) vs. other cities
lugano <- analysis_data %>% filter(City == "Lugano")
others <- analysis_data %>% filter(City != "Lugano")

cat(sprintf("Lugano (Ticino):        %.2f sales per 5,000 residents\n", lugano$Sales_per_5k)
```

```
Lugano (Ticino):        0.39 sales per 5,000 residents
```

```r
cat(sprintf("Other cities (average): %.2f sales per 5,000 residents\n", mean(others$Sales_per
```

```
Other cities (average): 1.28 sales per 5,000 residents
```

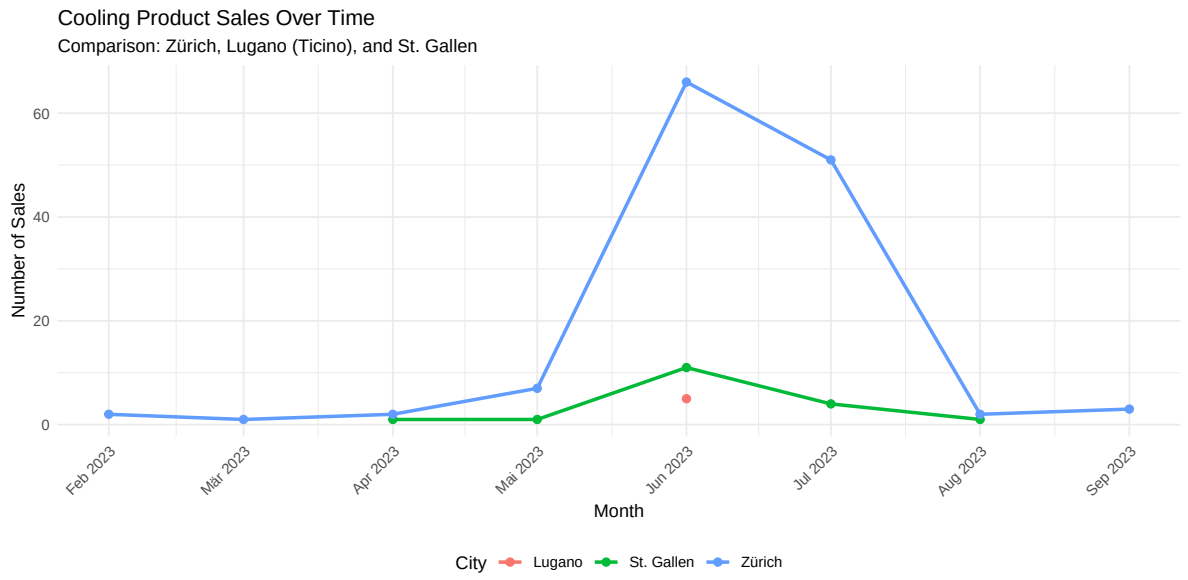
```r
cat(sprintf("Difference:              %.2f (%.1f%% %s)\n",
            lugano$Sales_per_5k - mean(others$Sales_per_5k),
            abs((lugano$Sales_per_5k / mean(others$Sales_per_5k) - 1) * 100),
            ifelse(lugano$Sales_per_5k > mean(others$Sales_per_5k), "higher", "lower")))
```

```
Difference:             -0.89 (69.3% lower)
```

7 6. Temporal Analysis

```r
# Sales by month
cooling_sales_filtered <- cooling_sales_filtered %>%
  mutate(
    date = as.Date(dateTime),
    month = format(date, "%Y-%m")
  )

monthly_sales <- cooling_sales_filtered %>%
  filter(cityName_clean %in% c("Zürich", "Lugano", "St. Gallen")) %>%
  count(month, cityName_clean) %>%
  mutate(month_date = as.Date(paste0(month, "-01")))

ggplot(monthly_sales, aes(x = month_date, y = n, color = cityName_clean)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month") +
  labs(
    title = "Cooling Product Sales Over Time",
    subtitle = "Comparison: Zürich, Lugano (Ticino), and St. Gallen",
    x = "Month",
    y = "Number of Sales",
    color = "City"
  ) +
```

```
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom"
  )
```



8 7. Conclusions

8.1 7.1 Key Findings

```
cat("KEY FINDINGS\n")
```

```
KEY FINDINGS
```

```
cat("============\n\n")
```

```
============
```

```
cat(sprintf("1. Correlation (Temperature - Sales): r = %.3f (p = %.4f)\n",
            cor_result$estimate, cor_result$p.value))
```

```
1. Correlation (Temperature - Sales): r = -0.479 (p = 0.2767)
```

```
cat(sprintf("\n2. Lugano (warmest city, 24.8°C): %.2f sales per 5k residents\n",
            analysis_data %>% filter(City == "Lugano") %>% pull(Sales_per_5k)))
```

```
2. Lugano (warmest city, 24.8°C): 0.39 sales per 5k residents
```

```
cat(sprintf("\n3. St. Gallen (coolest city, 19.6°C): %.2f sales per 5k residents\n",
            analysis_data %>% filter(City == "St. Gallen") %>% pull(Sales_per_5k)))
```

```
3. St. Gallen (coolest city, 19.6°C): 1.14 sales per 5k residents
```

```
cat(sprintf("\n4. Temperature difference: %.1f°C\n",
            max(analysis_data$Temperature) - min(analysis_data$Temperature)))
```

```
4. Temperature difference: 5.2°C
```

```
cat(sprintf("\n5. R²: %.3f (%.1f%% of variance explained by temperature)\n",
            summary(model)$r.squared,
            summary(model)$r.squared * 100))
```

```
5. R²: 0.230 (23.0% of variance explained by temperature)
```

8.2 7.2 Interpretation

Hypothesis: In Ticino, more cooling products are sold than in cooler regions.

Result: Hypothesis NOT confirmed (p >= 0.05)

The analysis shows a moderate negative correlation between average temperature and normalized sales.

Possible explanations:

1. **Temperature effect**: Higher temperatures lead to increased cooling demand
2. **Purchasing power differences**: Regional differences in income and buying behavior

3. **Urbanization**: City size and population density influence sales
4. **Seasonality**: Sales concentrate in summer months
5. **Data availability**: Limited sample (only 8 cities)

Created with: Quarto, R, tidyverse, ggplot2
Data sources: Digitec Galaxus Social Shopping Data, BFS Population Statistics, MeteoSwiss