

Weather's Cinematic Influence: Analyzing the Genre Preferences

Research Question



How do different weather conditions influence the genre preferences?

Introduction

‘*A rainy day calls for a cozy movie.*’

Reflects the belief that weather influences genre preferences.

What is “Cozy”? : a **subjective** perception

- a subjective perception related to warmth, comfort, relaxation, security

- Cozy Movie Genres:

Romance, Family Films → Lighthearted, comforting.

Less Cozy: Adventure, Action → High-energy, suspenseful. (*Zillmann, 1988*)

Introduction

Motivation:

A Potential Link Between Weather and Genre Preferences Mediated by Mood

- **Mood:** A sustained emotional state influencing decisions and experiences (APA, n.d.).

a potential relation between Weather and Genres through mood

- **Mood & Genre Preferences:**

Sad moods → social dramas or dark comedies / Happy moods → slapstick comedies or action films
(Greenwood, 2010).

- **Weather & Mood:**

Sunlight linked to positive mood and reduced tiredness (Denissen et al., 2008).

Winter is linked to a lower mood (Keller et al., 2005)

Introduction

Weather:

State of the atmosphere measured by factors like temperature, precipitation, pressure, humidity, and wind. (WMO)

Subjective Weather data are more suitable !

- Subjective terms like “warm” or “gloomy” reflect personal perception.

Challenges in Using Subjective Data

- Perceptions vary by experience, culture, and activity.
- Requires large-scale, time-sensitive surveys.
- Risk of memory errors and social biases.

Alternative: Objective Data

- Objective weather data is precise, consistent, and replicable.
- A certain amount of subjectiveness can be derived from objective data

Parameters:

Temperature

Precipitation height

Cloud Coverage

Wind Velocity

Introduction

Genre:

Describes similarities in films based on aesthetic, social, cultural, and psychological aspects. **(Bodenbjerg, 2015)**

Ideal Genre Data:

- Detailed viewer diaries tracking film choices, dominant genres, reasons for selection, and engagement duration.

Practical Challenges

- Requires extensive user participation and is time-intensive.
- Data privacy regulations of the data provider prevent access to personalized user information.

Chosen Metric: Watching Time per Genre

- Objective, reliable, and quantifiable measure of genre preference.
- Focuses on meaningful patterns in genre choices without personal data.

mood-related genres:

Comedy

Drama

Romance

Adventure

Thriller

non mood-related genres:

Knowledge

Research Question

How do different weather conditions influence the genre preferences?

Hypothesis:

Weather conditions have a an effect on mood-related genres

Methodology



Data Collection:

Weather Data

- Source: **Historical records from the German Weather Service (DWD)**
- Dataset: **Daily weather measures from Dahlem and Berlin Brandenburg weather station**
- **Period:** *01.01.2020 - 31.12.2023*

Deutscher Wetterdienst
Wetter und Klima aus einer Hand



Source: www.dwd.de

Method

```
STATIONS_ID,MESS_DATUM,QN_3,FX,FM,QN_4,RSK,RSKF,SDK,SHK_TAG,NM,VPM,PM,TMK,UPM,TXK,TNK,TGK,eor
403,20200101,-999,-999,-999,10,0.0,0,2.900,0,3.9,6.5,1022.90,1.8,94.00,5.0,-2.8,-6.3,eor
```

format: txt

Variables:

Temperature (°C)

Precipitation height (mm)

Cloud Cover (1/8)

Wind Velocity (m/s)

column name	description	uom	type	format
STATIONS_ID	Station ID		VARCHAR2	
MESS_DATUM	reference date		NUMBER	YYYYMMDD
QN_3	quality level of the following columns		NUMBER	numerical code
FX	daily maximum of windgust	m/s	NUMBER	9990.0
FM	daily mean of wind velocity	m/s	NUMBER	9990.0
QN_4	quality level of the following columns		NUMBER	numerical code
RSK	daily precipitation height	mm	NUMBER	9990.0
RSKF	precipitation form	numerical code	NUMBER	
SDK	daily sunshine duration	h	NUMBER	9990.0
SHK_TAG	daily snow depth	cm	NUMBER	9990.0
NM	daily mean of cloud cover	1/8	NUMBER	9990.0
VPM	daily mean of vapor pressure	hPa	NUMBER	9990.0
PM	daily mean of pressure	hPa	NUMBER	9990.0
TMK	daily mean of temperature	°C	NUMBER	9990.0
UPM	daily mean of relative humidity	%	NUMBER	9990.0
TXK	daily maximum of temperature at 2 m height	°C	NUMBER	9990.0
TNK	daily minimum of temperature at 2m height	°C	NUMBER	9990.0
TGK	daily minimum of air temperature at 5 cm above ground	°C	NUMBER	9990.0

Source: www.dwd.de

Data Collection

Data Collection:

Genre Data

- Source: **Users Watching Time from the FilmFriend platform**
- Users: **Users of public and university libraries in Germany, Austria, Switzerland, Belgium, Liechtenstein, Luxembourg and France.**
- Dataset: **Daily watching time per Zip Code per Genre.**
- Period: *01.01.2020 - 31.12.2023*



Source: <https://www.filmfriend.de/de/about-us>

Data Collection

Data Collection:

Genre Data

format: JSON

Variables:

Comedy

Drama

Romance

Adventure

Thriller

Knowledge

Watching Time per
Day, per Zip Code,
per Genre in
minutes



```

{
  "dateTime": "2020-01-01T00:00:00Z",
  "zipCode": "01445",
  "metricPerGenre": [
    {
      "id": 6,
      "englishName": "Book adaptation",
      "germanName": "Buchverfilmung",
      "metric": 278
    },
    {
      "id": 34,
      "englishName": "Drama",
      "germanName": "Drama",
      "metric": 1372
    },
    {
      "id": 35,
      "englishName": "Action",
      "germanName": "Action",
      "metric": 89
    }
  ],
  "metric": 1467
}

```

Methodology



Data Collection:

Zip Code Data

- Source: **Open source data from Open Data Informationsstelle Berlin (ODIS)**
- Dataset: **Zip Codes belong to the metropolitan area of Berlin (194 Zip Codes)**
- Format: CSV



Source: <https://odis-berlin.de/ueber-odis/>

Methodology

Data Preprocessing:

- **Weather Data:**

- Primary source: Berlin-Dahlem station, with **Wind Velocity** from Berlin-Brandenburg station.
- **Missing Values: 3 days** removed due to incomplete data.

- **Genre Data:**

- Converted raw data into **CSV format**.
- Filtered by **Berlin Zip Codes** to ensure location-specific relevance.
- **Daily Aggregation:** Viewing times were aggregated to **daily totals**.
- **Data Integration:**
- Weather and genre datasets were **merged by date** to allow for comparative analysis.

Methodology



Weather Data Labeling:

Defining weather labels based on the Terminology used by Weather Data providers:

German Weather Service

Temperature Level:

- 0 = **Cold** : Temperature less than 10 °C
- 1 = **Moderate** : Temperature between 10 and 20 °C
- 2 = **Warm** : Temperature more than 20 °C

Windy application

Precipitation Height Level:

- 0 = **Dry** : Precipitation Height less than 5 mm
- 1 = **Light Rain** : Precipitation Height between 5 and 10 mm
- 2 = **Heavy Rain** : Precipitation Height more than 10 mm

Heißer Tag

Ein Heißer Tag ist ein Tag, an dem das Maximum der Lufttemperatur ≥ 30 °C beträgt. Ein Heißer Tag wurde früher auch als Tropentag bezeichnet.

Source: <https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?nn=103346&lv2=101094&lv3=101162>

Sommertag

Ein Sommertag ist ein Tag, an dem das Maximum der Lufttemperatur ≥ 25 °C beträgt.

Source: <https://www.dwd.de/DE/service/lexikon/Functions/glossar.html;jsessionid=27D2CF374BB7FB1DB0E5EBEFF970DA35.live31092?lv2=102248&lv3=102522>

- + **Light rain** gives up to 2-4 mm (0.07-0.15 in) of precipitation;
- + **Moderate rain** gives 5-6 mm (0.19-0.23 in);
- + **Rain or strong rain** gives up about 15-20 mm (0.59-0.78 in);
- + **Rainfall** gives more than 30 mm (1.18 in).

Source: <https://windy.app/blog/how-do-we-measure-precipitation.html>

Methodology

Weather Data Labeling:

Defining weather labels based on the Terminology used by Weather Data providers:

U.S. national weather service

Wind Velocity Level:

- 0 = **Calm** : Wind Velocity less than 5 mph
- 1 = **Breezy** : Wind Velocity between 5 and 20 mph
- 2 = **Windy** : Wind Velocity more than 20 mph

Wind

The wind describes the prevailing direction from which the wind is blowing with speeds in miles per hour. The wind forecast is included in the first three periods of the zone forecast.

Sustained Wind	Speed Descriptive Term
0-5 mph	Light/ light and variable wind
5-10 mph/ 10-15 mph/ 10-20 mph	None
15-25 mph	Breezy (mild weather) Brisk or Blustery (cold weather)
20-30 mph	Windy
30-40 mph	Very Windy
40 mph or greater	Strong, dangerous, high, damaging (High Wind Warning Criteria)

Source: https://www.weather.gov/bgm/forecast_terms

Sky Condition

The sky condition describes the predominant/average sky cover based on percent of the sky covered by opaque (not transparent) clouds. If a high probability of precipitation (60% or greater) is expected, then the sky condition may be omitted since it is inferred from the precipitation forecast.

Sky Condition	Opaque Cloud Coverage
Clear/Sunny	1/8 or less
Mostly Clear/Mostly Sunny	1/8 to 3/8
Partly Cloudy/Partly Sunny	3/8 to 5/8
Mostly Cloudy	5/8 to 7/8
Cloudy	7/8 to 8/8

Source: https://www.weather.gov/bgm/forecast_terms

Cloud Cover Level:

- 0 = **Sunny** : Cloud Cover less than 4 Octal
- 1 = **Partly Cloudy** : Cloud Cover between 4 and 6 mph
- 2 = **Cloudy** : Cloud Cover more than 6 mph

Methodology

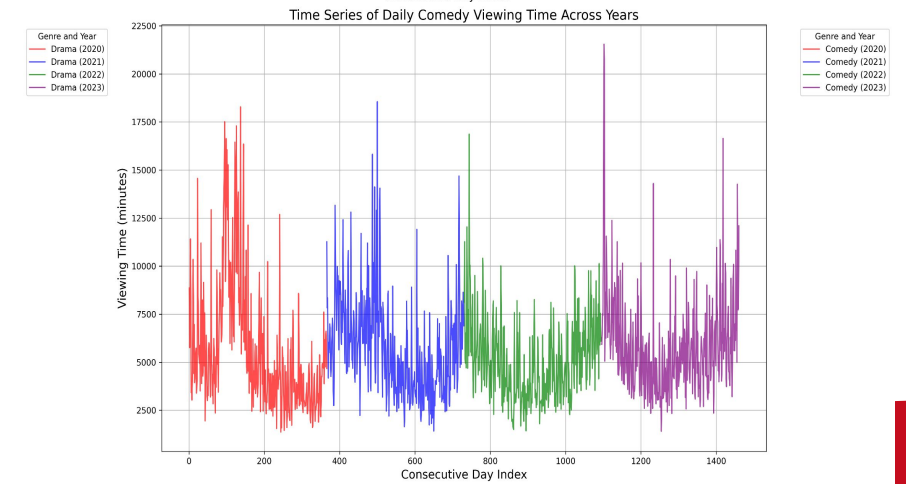
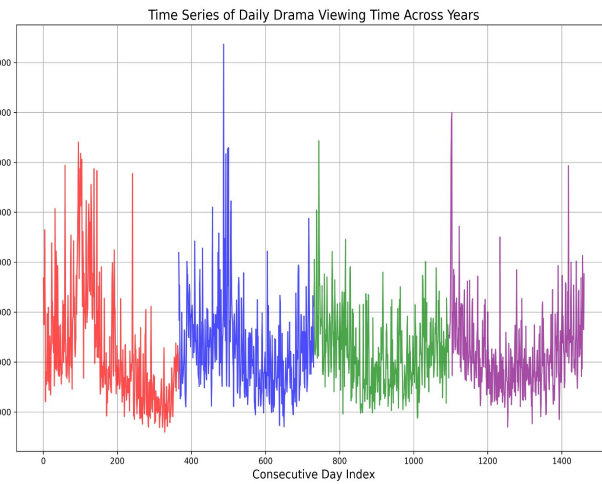
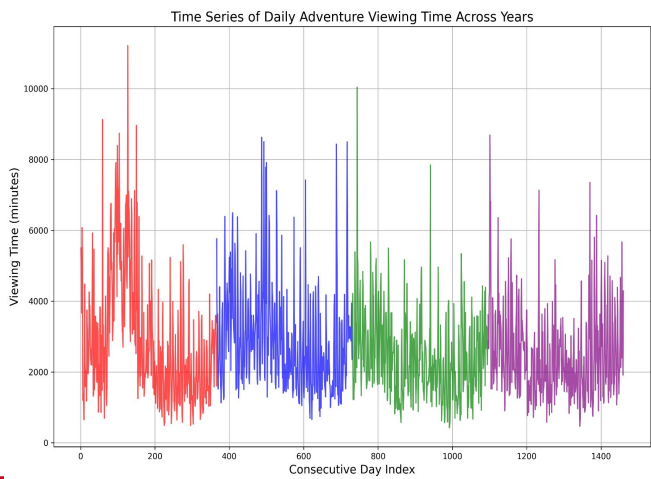
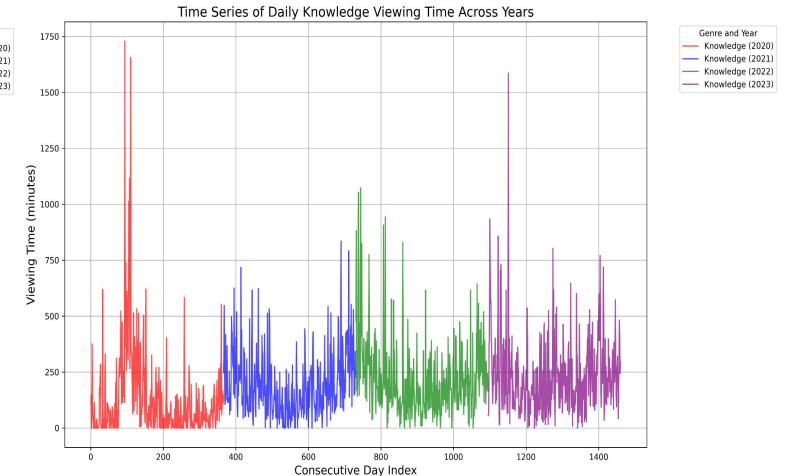
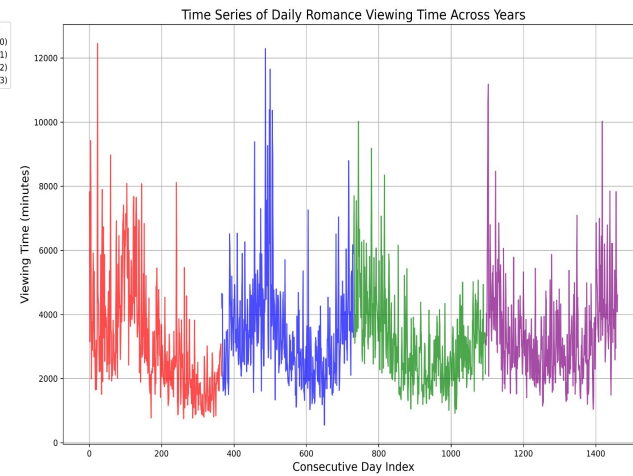
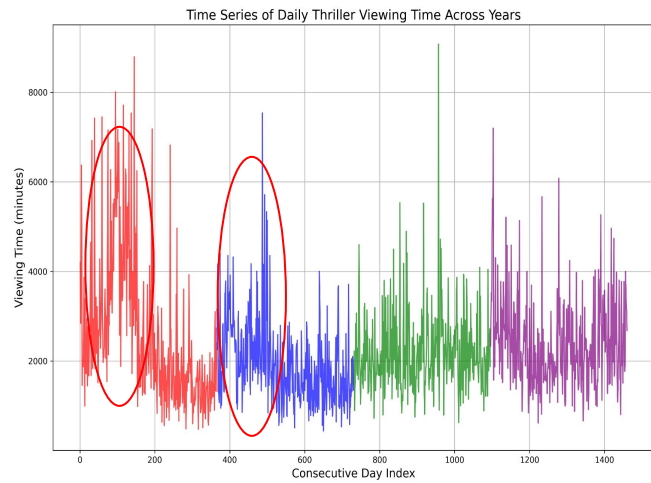


number of columns: 16

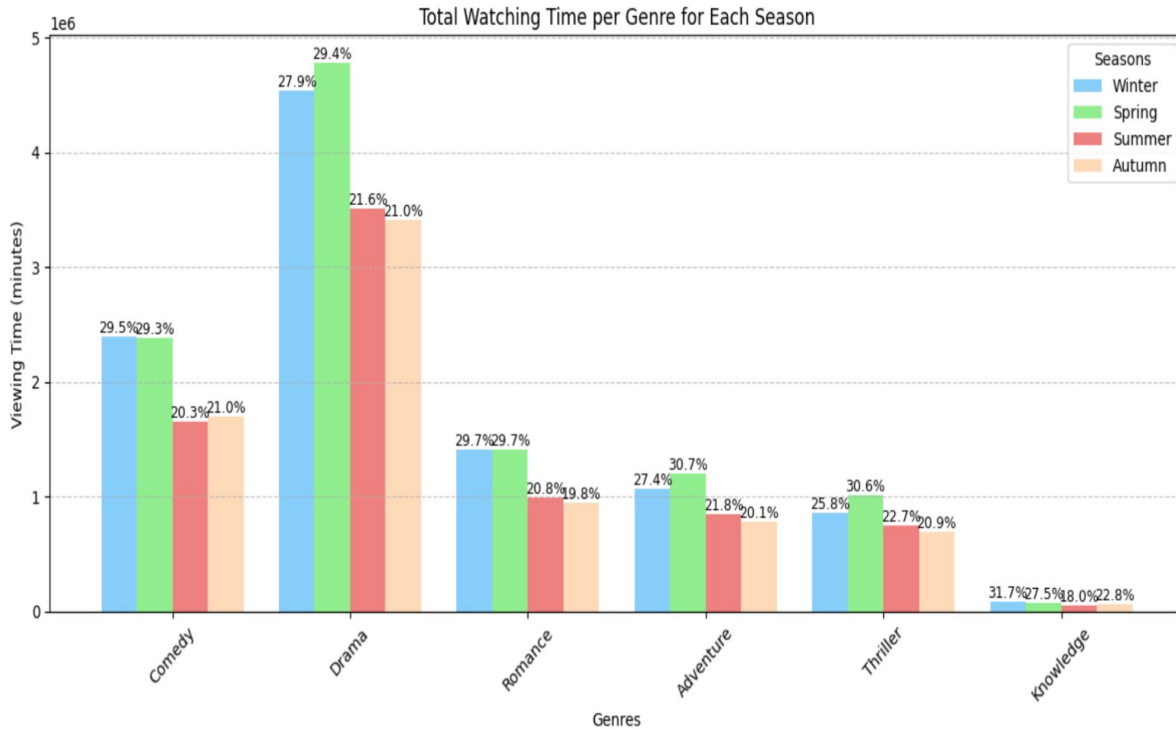
number of rows: 1457 (2020-2024 - 3 Days)

	dateTime	Temperature	Wind Velocity	Precipitation Height	Cloud Cover	Temperature Level	Wind Velocity Level	Precipitation Height Level	Cloud Cover Level	Comedy	Drama	Romance	Adventure	Thriller	Knowledge
0	2020-01-01	1.8	3.3	0.0	3.9	0	1	0	0	8857.333333	18400.000000	7817.000000	5502.666667	4203.000000	146.000000
1	2020-01-02	-0.3	2.9	0.0	1.8	0	1	0	0	5763.666667	13734.666667	3152.000000	4996.666667	2843.000000	0.000000
2	2020-01-03	4.3	6.5	4.4	6.6	0	1	0	2	7238.000000	14356.666667	5643.333333	3663.333333	4332.000000	0.000000
3	2020-01-04	4.1	7.7	3.0	4.8	0	1	0	1	11414.500000	23240.000000	9420.000000	6076.500000	6372.000000	375.000000
4	2020-01-05	1.4	3.9	0.0	4.1	0	1	0	1	8641.000000	17080.000000	6429.500000	5301.500000	4991.500000	114.000000
5	2020-01-06	4.7	4.3	0.0	7.2	0	1	0	2	3419.000000	6023.000000	1992.000000	1206.000000	1457.333333	48.000000
6	2020-01-07	2.3	3.5	0.2	5.3	0	1	0	1	3644.000000	8617.000000	3775.666667	1367.666667	2037.333333	48.666667
7	2020-01-08	6.2	5.8	1.9	7.9	0	1	0	2	3033.666667	9011.666667	2946.333333	1189.333333	1997.000000	0.000000
8	2020-01-09	8.1	4.4	5.3	8.0	0	1	1	2	4010.000000	7775.000000	2887.000000	653.666667	1871.333333	10.333333
9	2020-01-10	9.1	6.3	1.7	6.8	0	1	0	2	6706.000000	10309.333333	4460.000000	2290.666667	2730.666667	114.333333

Results



Results

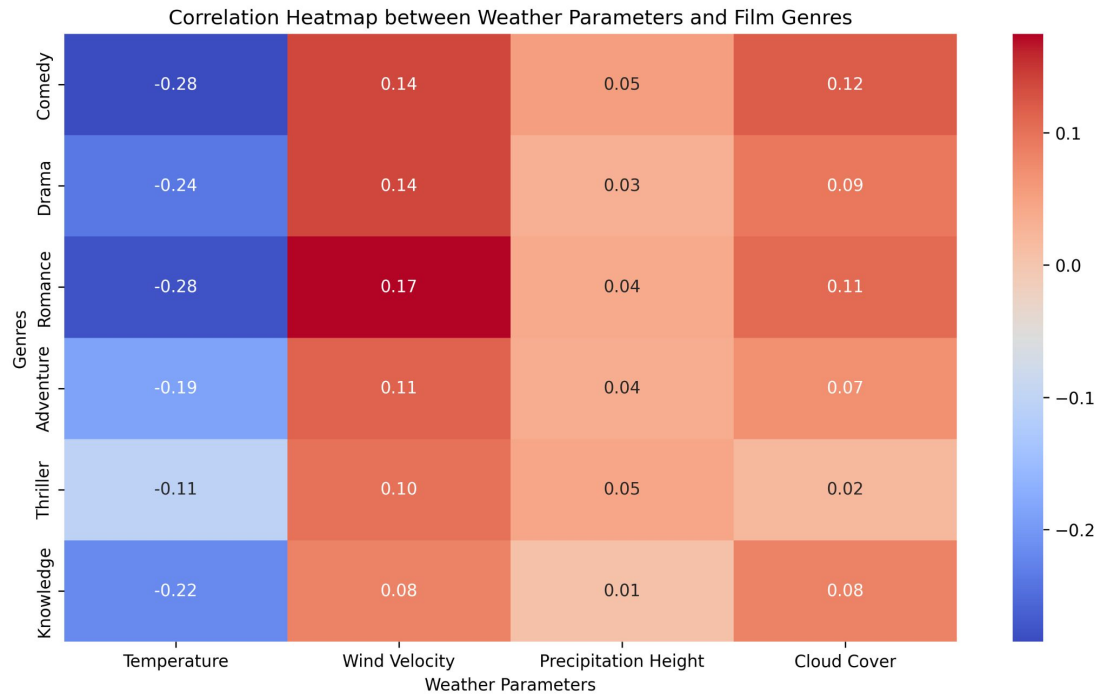


Seasonal Effect on Genre Viewing Time

- **Winter & Spring:** Similar viewing patterns observed.
- **Summer & Autumn:** Similar viewing patterns observed.
- **Seasonal Shifts:**
 - From **Spring to Summer:** Viewing time decreases by **7–10%** across all genres.
 - From **Autumn to Winter:** Viewing time increases by **6–9%** across all genres.

Seasonal variations follow a consistent pattern of cycles across all genres.

Correlation Analysis: 2020-2024

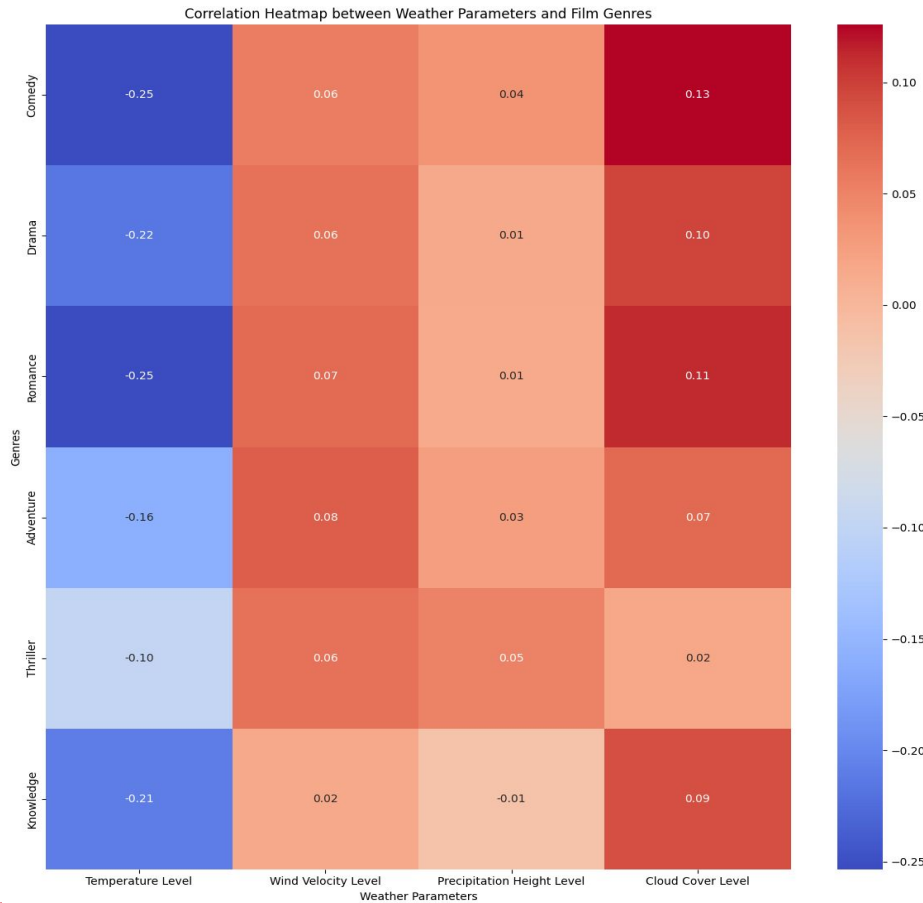


- **Temperature:** Negative correlation with genres but higher with **Comedy, Drama and Romance (-0.28 / -0.22)**
- **Wind Velocity:** Positive correlation with **Romance (0.17) & Comedy (0.14)**.
- **Cloud Cover:** Weak positive correlation with **Comedy (0.12) & Romance (0.10)**.
- **Precipitation:** No significant correlation with any genre.

Results



Correlation Analysis: With Weather Labels



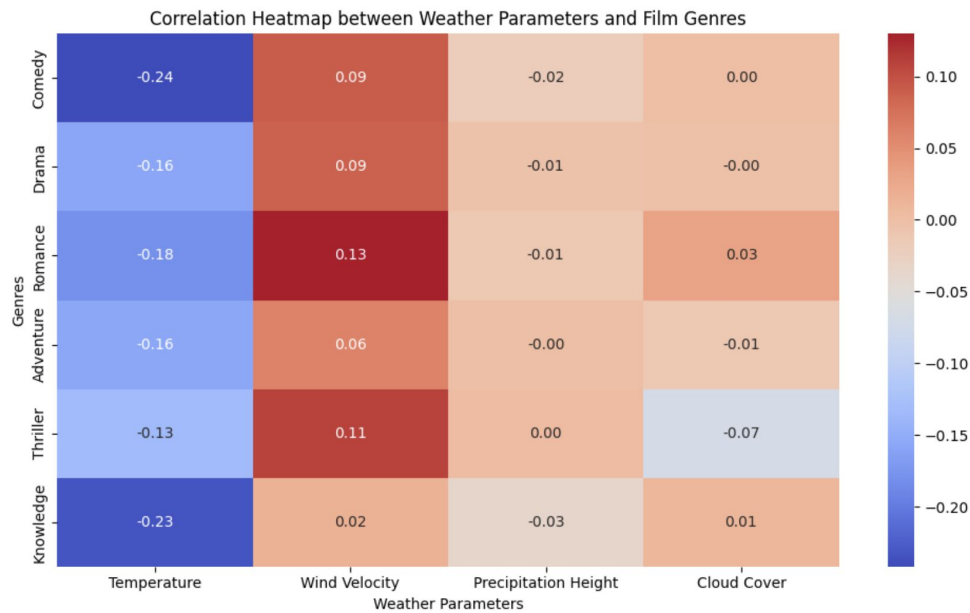
Weather categories capture similar trends but with diminished correlation strength.

Results

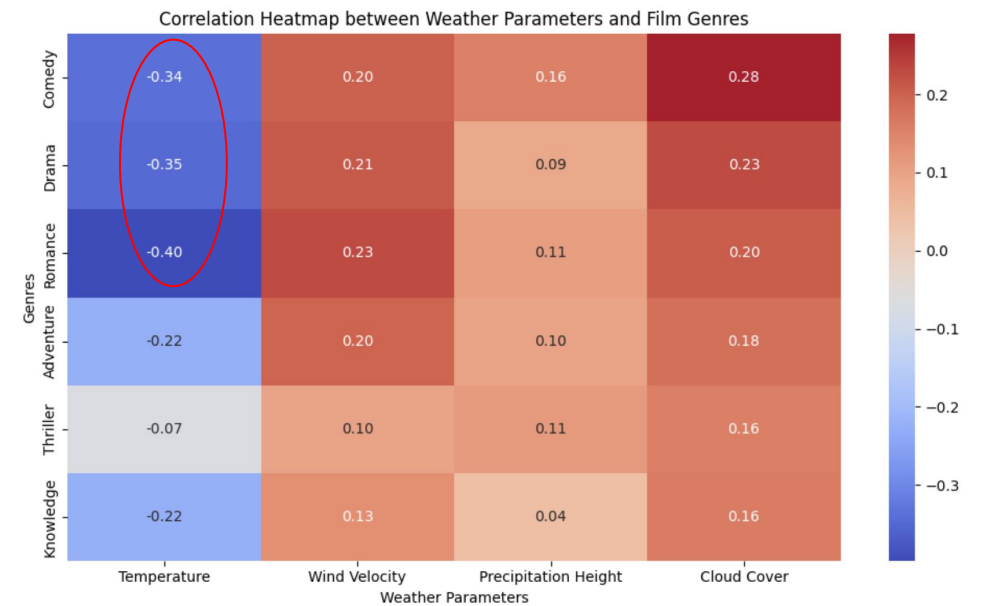


Correlation Analysis: Pandemic Effect

During Pandemic 2020-2021



After Pandemic 2022-2023



General Findings:

- **Seasonal and Weather-Related Viewing Patterns: Higher Viewing Times in Winter and Spring** compared to **Summer and Autumn**.
- **Colder Days:** Increased viewing, especially for **Drama, Romance, and Comedy**.
- **Windy & Cloudy Days:** Similar effect, but less pronounced.
- **Precipitation Levels:** No significant influence on genre preferences.

Overall Correlations: Weak and not strong enough to fully explain the relationship.

Challenges and Risks



- **Intra-City Weather Variability:** Berlin's size may lead to varying weather conditions across different districts at the same time.
- **Lack of Viewing Context:** Data doesn't reveal **why** a genre was watched (e.g., actors, ratings, platform layout, or series completion).
- **Genre Overlap:** Many films and episodes span **multiple genres**, complicating genre-specific analysis.
- **Geographic Limitation:** Findings are based solely on **Berlin weather data** and may not generalize to other regions.

Questions and Feedback