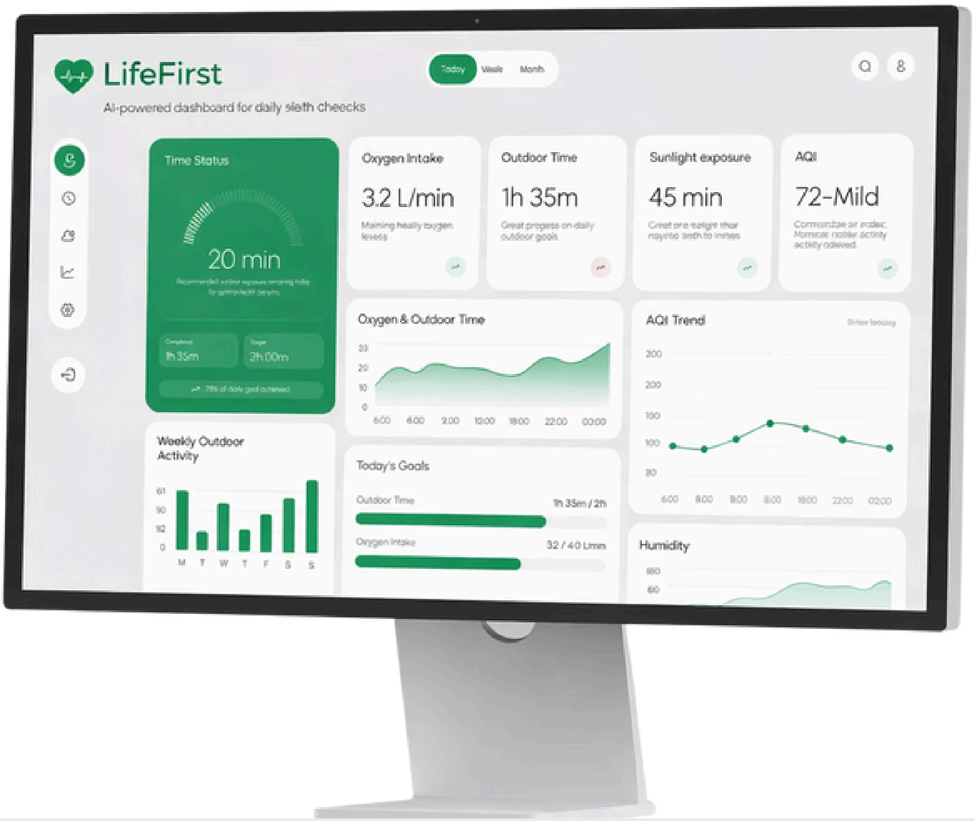


LifeFirst

AI-powered dashboard for daily sleth cheecks

Time Status

20 min

Recommended outdoor exposure remaining today for optimal health benefits

Completed 1h 35m

Target 2h 00m

78% of daily goal achieved

Oxygen Intake

3.2 L/min

Maining healty oxygen levess

Outdoor Time

1h 35m

Great progess on daily outdoor goals

Sunlight exposure

45 min

Creat ore selight thar nayrse sleth to imtes

AQI

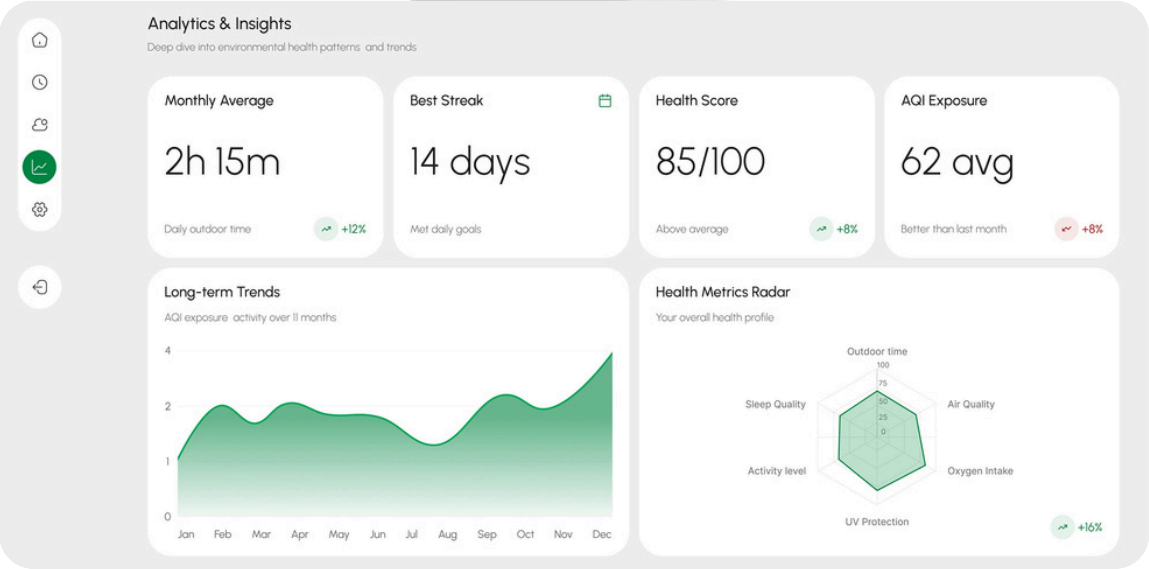
72-Mild

Comontze air erdec Momcoc rioder actity actity odeved

Today's Goals

Outdoor Time 1h 35m / 2h

Oxygen Intake 32 / 4.0 L/min



Analytics & Insights

Deep dive into environmental health patterns and trends

Monthly Average

2h 15m

Daily outdoor time +12%

Best Streak

14 days

Met daily goals

Health Score

85/100

Above average +8%

AQI Exposure

62 avg

Better than last month +8%

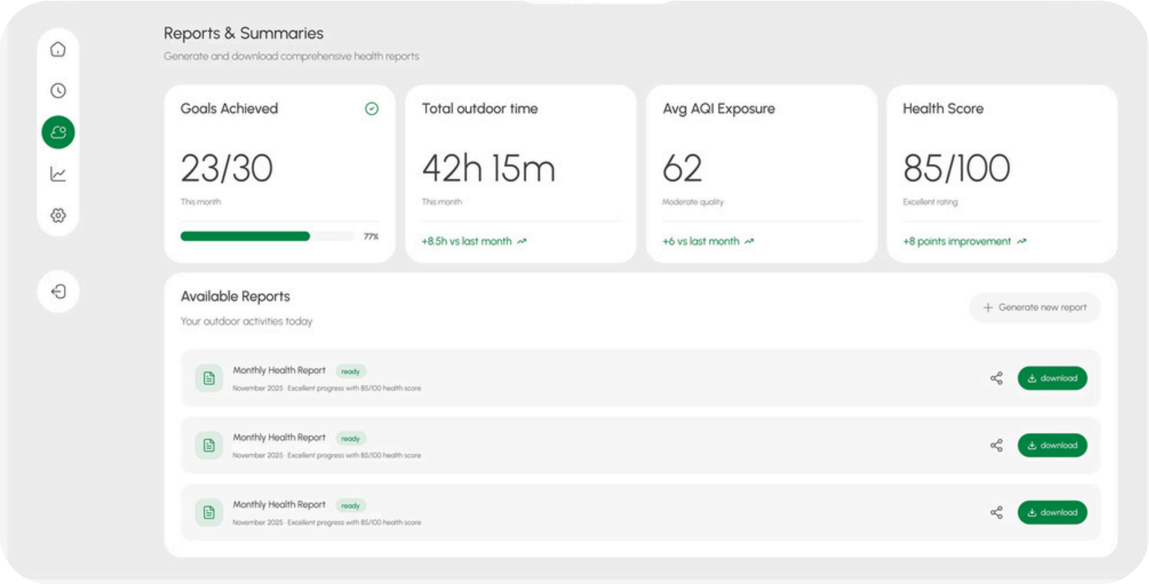
Long-term Trends

AQI exposure activity over 11 months

Health Metrics Radar

Your overall health profile

+16%



Reports & Summaries

Generate and download comprehensive health reports

Goals Achieved

23/30

This month

77%

Total outdoor time

42h 15m

This month

+8.5h vs last month

Avg AQI Exposure

62

Moderate quality

+6 vs last month

Health Score

85/100

Excellent rating

+8 points improvement

Available Reports

Your outdoor activities today

Monthly Health Report ready

November 2025 · Excellent progress with 85/100 health score

Monthly Health Report ready

November 2025 · Excellent progress with 85/100 health score

Monthly Health Report ready

November 2025 · Excellent progress with 85/100 health score

LifeFirst | AI powered Health analytics dashboard

Product&Data Case Study
Urban Health · Lifestyle Analytics · Decision Support
Data Sources: Kaggle (Healthy Lifestyle Cities Report 2021), open-source GitHub notebooks
Tools: Python, pandas, NumPy, scikit-learn, Plotly, Matplotlib, Seaborn
Author: Shashwat Chauhan
Institution: BITS Pilani, Goa Campus
Date: JUNE 2025

Overview

Large volumes of public health and lifestyle data are available through platforms such as Kaggle and GitHub. However, most of this data remains confined to notebooks and static reports, limiting its usefulness for real decision-making.

Users commonly face the following challenges:

- Health indicators are spread across disconnected dimensions
- Raw metrics lack context, prioritization, and comparability
- Reports describe data but rarely support interpretation or action LifeFirst reframes this.

How can open urban health datasets be transformed into a structured analytics pipeline that supports comparison, interpretation, and insight generation?

LifeFirst is an AI-powered dashboard that presents health as a **system-level outcome** shaped by lifestyle, environment, and socio-economic factors. The product prioritizes clarity over complexity, converting dense datasets into signals that are easy to read, compare, and reason about.

1. Product Context

Health outcomes are influenced by where people live, how they work, what they can afford, and how their environment behaves over time. LifeFirst treats these dimensions as interconnected rather than isolated statistics.

The dashboard is designed for:

- Individuals seeking awareness of lifestyle health factors
- Researchers and analysts exploring urban health patterns
- Product and policy teams comparing cities at scale

The core design objective is simple: **reduce cognitive load while preserving real-world nuance**.

2. Problem Framing

At the outset, no assumptions are made about data quality or completeness. Open datasets are treated as imperfect but informative signals.

This case study documents how an openly available dataset was operationalized into a **production-style analytics workflow** within LifeFirst, focusing on interpretability, reproducibility, and decision support.

3. Data Foundation

LifeFirst is built on open and reproducible data. The primary dataset used is the **Healthy Lifestyle Cities Report 2021**, sourced from Kaggle and widely referenced in open-source analyses.

The dataset covers 43 major cities and captures multiple dimensions commonly associated with quality of life. Key feature groups include:

- **Environmental exposure:** sunshine hours, pollution levels
- **Lifestyle accessibility:** cost of water, gym memberships
- **Health outcomes:** obesity levels, life expectancy
- **Behavioral proxies:** outdoor activity indicators, annual working hours

This structure aligns with contemporary public health research, which treats health as a socio-environmental outcome rather than a purely biological one.

4. Data Ingestion

The ingestion layer in LifeFirst is intentionally lightweight and reproducible. Raw datasets are treated as immutable inputs so the pipeline can be rerun as new versions of the data become available.

```
import pandas as pd
import numpy as np

train = pd.read_csv('healthy_lifestyle_city_2021.csv')
```

5. Data Cleaning and Transformation

Open datasets frequently encode numeric values as strings due to currency symbols, percentages, or placeholders. These values must be standardized before analysis.

```
train['Cost of a bottle of water (City)']      = (
    train['Cost of a bottle of water (City)'].str.replace('£',          '').astype(float)
)

train['Obesity levels (Country)']     = (
    train['Obesity levels (Country)'].str.replace('%',        '').astype(float)
)

train['Cost of a monthly gym membership (City)']  = (
    train['Cost of a monthly gym membership (City)'].str.replace('£',
'').astype(float)
)

train = train.replace('-', np.nan)
train = train.drop(['Rank'], axis=1)
```

This step ensures numeric consistency across features and mirrors common preprocessing practices in applied data science.

6. Handling Missing Data

Missing values were limited and addressed using **domain-aware completion** rather than blind statistical imputation. Where possible, public benchmarks and government sources were used to preserve realistic distributions.

This approach maintains dataset completeness without distorting global patterns.

7. Feature Validation

After transformation, the dataset contains 44 city records and 11 usable numerical features. Each feature was reviewed to ensure it meaningfully contributes to lifestyle or health interpretation within LifeFirst.

8. Exploratory Analysis

Exploration in LifeFirst focuses on understanding **distributions and variance**, Key

observations:

- Most indicators are not normally distributed
- Several cities appear as meaningful outliers
- Cost and pollution metrics vary more widely than biological indicators

These findings influenced both modeling decisions and dashboard visual design.

9. Visual Analytics Pipeline

Visualizations are designed around **specific user questions**, not aesthetics.

- Maps surface regional clustering effects
- Histograms support distribution awareness
- Box plots highlight spread and outliers

```
import plotly.express as px

px.scatter_mapbox(
    train,
    lat='latitude',
    lon='longitude',
    size='Happiness levels (Country)',
    color='Pollution (Index score) (City)',
    zoom=1
)

px.histogram(train, x='Sunshine hours (City)')
px.box(train, y='Pollution (Index score) (City)')
```

10. Geographic Mapping

Location provides essential context for environmental and lifestyle indicators. Geographic views allow users to identify patterns that are difficult to detect in tabular form.

```
from geopandas.tools import geocode

location = geocode(train['City'], provider='nominatim', user_agent='lifefirst')
train['longitude'] = location.geometry.x
train['latitude'] = location.geometry.y
```

11. Statistical Relationships

Happiness score is treated as a composite proxy for perceived quality of life.

Observed patterns include:

- Positive correlation between happiness, life expectancy, and cost of living
- Negative correlation with pollution levels and long working hours

These relationships informed metric prioritization within the dashboard.

12. Predictive Modeling

Predictive models were used to test whether lifestyle and environmental features jointly explain variation in happiness levels.

ElasticNet was selected for its balance between regularization and interpretability, aligning with LifeFirst’s emphasis on explainable insights.

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import ElasticNet

X = train.drop('Happiness levels (Country)',    axis=1)
y = train['Happiness levels (Country)']

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.1, random_state=42
)

model = ElasticNet(alpha=0.45, l1_ratio=0.00002)
model.fit(X_train, y_train)
```

13. Clustering for City Typologies

To surface higher-level patterns, cities were grouped into lifestyle archetypes using unsupervised clustering.

```
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans

scaler = MinMaxScaler()
scaled = scaler.fit_transform(train.drop('City', axis=1))

kmeans = KMeans(n_clusters=6, random_state=42)
train['cluster'] = kmeans.fit_predict(scaled)
```

These clusters enable features such as city benchmarking and similarity-based comparisons within LifeFirst.

14. Product Integration

All analytical outputs are surfaced through a dashboard designed for comparison, exploration, and insight generation rather than raw data inspection.

Model outputs inform:

- City comparison views
- Lifestyle archetype grouping
- Metric prioritization across dashboard sections

15. Final Product Takeaway

LifeFirst demonstrates how open urban health data can be transformed into a **decision-support product** rather than a static report.

By emphasizing interpretability, system-level thinking, and thoughtful visual design, LifeFirst turns complex lifestyle data into insights that are accessible, comparable, and actionable.

Project: LifeFirst — Health Analytics Dashboard
Author: Shashwat Chauhan
Institution: BITS Pilani, Goa Campus