

Memoria técnica [EDA-Nomadismo Digital]

AUTORES:

- **Juan F. Cía** — [LinkedIn](#) · [GitHub](#)
- **Daniela A. Aguirre** — [LinkedIn](#) · [GitHub](#)
- **Alejandro Balaguer** — [LinkedIn](#) ·

Fecha de entrega:

09/01/2026

Repositorio:

EDA-Nomadismo-Digital

1. Resumen ejecutivo:

Pregunta de investigación

A nivel global, ¿qué características económicas y de coste de vida diferencian a los destinos mejor posicionados para el nomadismo digital frente a los menos atractivos?

Objetivo general

Realizar un Análisis Exploratorio de Datos (EDA) para estudiar la relación entre el posicionamiento/ranking de destinos para nómadas digitales y variables económicas/de coste de vida (según disponibilidad en el dataset), con el fin de identificar qué factores explican mejor las diferencias entre destinos más y menos atractivos, y si estos factores influyen también al buscar calidad de vida.

Este proyecto realiza un Análisis Exploratorio de Datos (EDA) a nivel global para identificar qué variables económicas y de coste de vida diferencian a los destinos mejor posicionados para el nomadismo digital frente a los menos atractivos. El análisis se basa en un dataset de destinos internacionales que incluye un ranking de atractivo y métricas relacionadas con coste de vivienda, precios de bienes/servicios esenciales y, cuando aplica, indicadores de ingresos o salario.

El trabajo consistió en la revisión y limpieza de los datasets [Numbeo/Kaggle](#) | [Circleloop](#) y [Movingto](#) (tratamiento de valores faltantes, verificación de rangos y consistencia de unidades), la generación de estadísticos descriptivos y comparativos por grupos de ranking (por ejemplo: top vs bottom), y el estudio de relaciones entre variables mediante correlaciones y visualizaciones (distribuciones, boxplots, heatmaps de correlación y scatter y regplots multidimensionales.). Además, se exploraron posibles casos atípicos ("outliers") que combinan coste moderado con alta valoración.

2. Contexto y objetivos

El nomadismo digital ha crecido como alternativa laboral y de estilo de vida, pero la elección de destino suele depender de factores económicos (coste de vida, vivienda, precios esenciales) además de la calidad de vida. Este proyecto analiza, a nivel global, qué variables económicas y de coste de vida están asociadas con un mejor posicionamiento de los destinos para nómadas digitales.

Objetivo general: identificar patrones y diferencias entre destinos mejor y peor rankeados mediante un Análisis Exploratorio de Datos (EDA), evaluando el peso relativo de variables como vivienda, coste total, salarios y precios de bienes/servicios esenciales.

Objetivos específicos:

- Comparar costes entre grupos (top vs bottom del ranking).
 - Evaluar el papel del coste de vivienda como factor diferenciador.
 - Explorar coherencia entre salarios e índices de precios (poder adquisitivo).
 - Detectar outliers atractivos (alta valoración con coste moderado).
 - Identificar qué rubros de gasto impactan más en la viabilidad.
-

3. Requisitos:

Requisitos funcionales

- Importar y explorar los datasets (estructura, tipos de datos, duplicados).
- Limpiar datos (valores nulos, formatos, unidades si aplica).
- Generar estadísticos descriptivos y comparaciones por grupos.
- Crear visualizaciones para soportar conclusiones (distribuciones, boxplots, heatmaps de correlación y scatter y regplots multidimensionales.)
- Identificar outliers y documentar casos relevantes.

Requisitos no funcionales

- Reproducibilidad: ejecución mediante notebook/script con dependencias definidas.
 - Claridad: resultados explicados con tablas y gráficas interpretables.
 - Trazabilidad: cambios y avances gestionados en GitHub.
-

4. Diseño y arquitectura:

El proyecto se estructura como un pipeline de análisis exploratorio: 4.1. **Carga de datos** desde el archivo fuente del dataset: -1) Clonar el repositorio git clone <https://github.com//EDA-Nomadismo-Digital.git> cd EDA-Nomadismo-Digital

4.2. Crear y activar entorno virtual -python -m venv .venv -source .venv/bin/activate # Linux/Mac -
-.venv\Scripts\activate # Windows

4.3. Instalar dependencias -pip install -r requirements.txt

4.5. Levantar Jupyter (opcional, si el análisis está en notebooks) -jupyter lab

4.6. **Preparación:** limpieza, tratamiento de nulos, validaciones y variables derivadas.

4.7. **EDA:** análisis descriptivo, segmentación por ranking y exploración de relaciones entre variables.

4.8. **Comunicación:** generación de visualizaciones y conclusiones (notebook + presentación). **Estructura del repositorio (referencial):**

- **data/**: datasets originales utilizados:

Dataset	Registros	Variables	Fuente
Cost of Living	4.742 ciudades	65 variables	Numbeo/Kaggle
Circleloop Index	85 países	10 variables	Circleloop
Movingto Index	40 países	10 variables	Movingto

- **requirements.txt**: Entorno y dependencias: El proyecto se desarrolló en Python y las dependencias del entorno están definidas en **requirements.txt**, lo que permite reproducir el análisis instalando las librerías necesarias. Aunque el archivo incluye dependencias transitivas, para el EDA se utilizaron principalmente librerías de análisis y visualización de datos (p. ej., pandas/numpy y matplotlib/seaborn, además de Jupyter para la ejecución de notebooks).

5. Implementación:

El EDA se desarrolló en notebooks de Jupyter dentro de la carpeta **notebooks/**. El flujo de trabajo incluye: carga de los datasets, limpieza, análisis descriptivo, comparaciones por grupos del ranking, variables agregadas, visualizaciones y detección de outliers, EDA-bivariante, EDA-univariante, main-analysis, planteamiento-EDA.

Archivos principales:

- **notebooks/[NOMBRE].ipynb**: (Cuando ya tengamos notebook final: renombrar a algo neutro)
- **Notebooks por fase**
 - **ETL / Preparación de datos**: notebooks encargados de limpiar, homogeneizar e integrar variables de coste de vida e índices.
 - **Variables agregadas**: notebooks donde se crean variables derivadas y métricas para el análisis.
 - **EDA univariante y bivariante**: notebooks con estadística descriptiva, comparaciones por grupos y relaciones entre variables.
 - **Notebook principal**: consolidación de resultados, visualizaciones finales y conclusiones.

Nota: aunque algunos archivos mantienen nombres individuales, el contenido final integra el trabajo del equipo.

-Datasets utilizados:

- El análisis se realizó sobre el dataset descrito en el punto 1 y 4, tras un proceso de limpieza y preparación (ETL) documentado en los notebooks correspondientes.
- **img/**: gráficas exportadas usadas en la presentación y documentación (pendiente)
- **docs/memoria-tecnica.md**: memoria técnica del proyecto (se añade al finalizar)

6. Pruebas:

Al tratarse de un análisis exploratorio, las "pruebas" se enfocaron en validaciones de calidad del dato:

- Revisión de valores nulos y estrategia de tratamiento (eliminación/imputación según variable).
 - Búsqueda de duplicados y verificación de claves.
 - Validación de rangos razonables (por ejemplo, costes y salarios no negativos).
 - Comprobación de coherencia de unidades/monedas cuando fue necesario, se hizo una estandarización de monedas a USD.
 - Revisión de outliers para distinguir errores de registro vs. casos reales.
-

7. Resultados:

Los resultados del EDA se presentan organizados según las preguntas iniciales del proyecto. Para ello, se compararon destinos por grupos de ranking (por ejemplo, destinos mejor posicionados frente a peor posicionados), se analizaron distribuciones y relaciones entre variables económicas y se revisaron valores atípicos (outliers), se detectaron ciudades con problemas de gentrificación.

7.1 ¿Son realmente más baratos los destinos mejor rankeados?

La comparación por grupos sugiere que existen diferencias en el coste de vida entre destinos mejor y peor posicionados, aunque dichas diferencias no son uniformes para todos los rubros. El análisis establece los cuartiles (25% y 75%), la mediana y el percentil 90

Evidencia: [Figura/Tabla: comparación de coste total por grupos de ranking].

7.2 ¿El coste de vivienda es el principal factor diferenciador?

Entre las variables analizadas, el coste de vivienda destacada por la hipótesis 4: El coste en vivienda no es relevante. Comer fuera o el gasto en la cesta sí. lo que indica que este componente tiene un peso importante en la viabilidad económica del destino.

Evidencia: [Figura: boxplot de alquiler por grupos] y/o [Tabla: ranking de variables por correlación].

7.3 ¿Existe coherencia entre salarios locales y precios esenciales?

Al contrastar variables de ingresos con precios de productos/servicios esenciales, se observa que los ciudadanos del 73,7% de las ciudades ingresan un salario promedio superior al gasto mensual y los ciudadanos del 26,3% de las ciudades ingresan un salario promedio inferior al gasto mensual. Esto sugiere que el ranking no depende únicamente del nivel salarial, sino también de la estructura de costes y de la relación entre ingresos y gasto básico. Análisis del gasto porcentual del salario promedio en vivienda:

Situación óptima (gasto < 30%): 1591 ciudades (33.6%) Situación mejorables (gasto entre 30-50%): 2014 ciudades (42.5%) Situación arriesgada (gastos > 50%): 1137 ciudades (24.0%) **Evidencia:** [Figura: scatter salario vs coste] / [Figura: índices de poder adquisitivo].

7.4 ¿Hay "outliers" atractivos?

Los diagramas de caja muestran valores atípicos en algunas variables, como el coste de vida mensual, el coste de productos básicos y el alquiler de vivienda. Sin embargo, en el caso de las variables más relevantes para los nómadas digitales, como la velocidad de conexión a internet y el índice de bienestar, no se identificaron outliers significativos. Además, en la validación de la hipótesis 7, se identificaron países que son outliers positivos en términos de menor coste mensual y mejor puntuación en el ranking de destinos para nómadas

digitales. Los tres principales países en esta categoría son Rumanía, Hungría y Lituania. Como “outliers negativos” (casos extremos de presión de vivienda): “Se identifican ciudades con valores extremos de esfuerzo de vivienda, que podrían distorsionar comparativas y requieren interpretación contextual.”

7.5 ¿Qué tipo de gasto pesa más en la viabilidad del nomadismo digital?

El gasto en alimentación y la cesta básica tiene un mayor peso en la viabilidad del nomadismo digital, en comparación con el coste de la vivienda. Por otro lado, también se destaca que la relación calidad-precio de los productos y servicios (hipótesis 5) tiene una correlación positiva con la elección de los destinos, lo que indica que los nómadas digitales valoran más los lugares donde pueden obtener mejores condiciones por un menor coste mensual. **Evidencia:** [Tabla/Figura: contribución o comparación de rubros].

7.6 Resumen de respuestas

1. ****Países con mejor equilibrio gasto/ score (Hipotesis 7)** 🙄 *Sí — En el top 10 (menor monthly_nomad_cost y mayor digital_nomad_score) destacan Rumanía, Hungría, Lituania, Tailandia, Letonia, Chile, Estonia, Brasil, Croacia y Rusia.
 2. **Vivienda como factor principal:** El coste en vivienda no es relevante. Comer fuera o el gasto en la cesta sí
 3. **Coherencia salarios-precios:** Media: Aunque en el 73,7% de ciudades el salario promedio supera el gasto mensual, en un 26,3% no lo hace, lo que indica que la relación ingresos-costes no es consistente y el ranking también depende de la estructura de costos.
 4. **Outliers atractivos:** Los tres principales países en esta categoría son Rumanía, Hungría y Lituania.
 5. **Gasto más determinante:** El gasto en alimentación y la cesta básica es el más determinante.
-

8. Despliegue y uso

8.1 Requisitos: -Python: 3.x -Librerías: pandas, numpy, matplotlib/seaborn. -(Opcional): Sistema operativo: Windows/macOS/Linux.

8.2 Estructura del proyecto: -/docs/ → documentación del proyecto (memoria técnica, etc.) -/src/ → directorio principal del trabajo -/src/data/ → datasets/archivos de datos usados en el análisis -/src/img/ → imágenes (gráficas, capturas, recursos) -/src/notebooks/ → notebooks del proyecto (ETL, EDA y análisis final) - main_analisis_juanfcia_v1.ipynb → análisis principal -ETL_* → notebooks de ETL -EDA-* → notebooks de EDA (uni/bivariante) -variables_agregadas_* → preparación/agregación de variables -/src/utils/ → funciones auxiliares y utilidades reutilizables

Crear entorno (opcional pero recomendado): -python -m venv venv -activar entorno

Instalar dependencias: -pip install -r requirements.txt -Instalar librerías necesarias indicadas en el notebook

8.3 Ejecución / Reproducción del análisis: -Abrir el notebook principal: src/notebooks/main_analisis.ipynb

-Los resultados (tablas y visualizaciones) se generan en el propio notebook.

8.4 Uso: ¿qué puede hacer el usuario con el proyecto? -Consultar un ranking de destinos para nómadas digitales a nivel país (y, cuando aplique, ciudad) en función de coste mensual estimado y puntuación/score de atractivo para nómadas digitales. [The_Bridge] EDA - Nomadismo digital

-Analizar relaciones entre variables clave (coste de vida, poder adquisitivo, conectividad, bienestar, visados, etc.) mediante comparativas y correlaciones para entender qué factores están más asociados al score final.

-Validar hipótesis del análisis (p. ej., peso de la vivienda frente a otros costes, coherencia salarios-precios, papel de la conectividad, detección de outliers y países "mejor equilibrio coste/score").

-Comparar países/ciudades con filtros por continente o por métricas (coste mensual, alquiler, cesta básica, velocidad de internet, índice de felicidad, seguridad, facilidad de visa), facilitando la toma de decisiones según prioridades del usuario.

8.5 Limitaciones y consideraciones: -Dependencia de fuentes externas: los valores provienen de datasets de terceros y pueden contener errores, sesgos o desactualización respecto a la situación real del país/ciudad.

-Resultados agregados: el análisis trabaja con tendencias globales/medias; no sustituye una investigación local (barrios, estacionalidad, seguridad real por zona, requisitos de visado vigentes, fiscalidad específica, etc.).

-Comparabilidad limitada entre países/ciudades: diferencias en metodología de recopilación, divisa, inflación o coste real para perfiles concretos pueden afectar la interpretación de "barato/caro".

-Correlación \neq causalidad: relaciones encontradas no implican causa-efecto; deben usarse como guía para priorizar variables, no como prueba definitiva.

-Cobertura incompleta: no todos los países/ciudades aparecen en todas las fuentes, por lo que algunos rankings o comparativas pueden tener muestras diferentes.

9. Gestión del proyecto:

Alcance y objetivo: se desarrolló un EDA para identificar qué características económicas y de coste de vida diferencian los destinos mejor posicionados para el nomadismo digital frente a los menos atractivos.

-Enfoque por hipótesis: el análisis se planificó y ejecutó alrededor de 9 hipótesis (3 primarias y 6 secundarias) sobre coste, salarios/poder adquisitivo, conectividad, vivienda, "más por menos", bienestar, outliers, seguridad y facilidad de visado.

-Fases de trabajo (workflow)

-Contexto e hipótesis (definición de preguntas y variables objetivo).

-Entendimiento de datos: revisión de datasets y variables (Cost of Living + índices de nomadismo).

-EDA univariante: distribuciones, percentiles y outliers en variables relevantes (vivienda, alimentación, transporte, salarios, etc.).

-Preparación e integración: merge de datasets para análisis bivalente y validación de hipótesis con variables agregadas y scores.

-EDA bivalente + validación: matriz de correlación y visualizaciones tipo scatter/regresión para contrastar hipótesis vs. digital_nomad_score.

-Síntesis y comunicación: resumen de correlaciones y estado (☒/✗) de cada hipótesis.

-Organización del equipo y entregables: trabajo en equipo (Team 4: Daniela Aguirre, Alejandro Balaguer y Juan F. Cía) y entrega en formato presentación con resultados, visualizaciones y conclusiones.

10. Conclusiones y trabajo futuro:

-La conectividad es el factor más asociado al score: broadband_speed_mbps presenta la correlación más alta con el score, seguida por bienestar (happiness_index) y variables de gasto (comidas fuera y gasto mensual).

-No gana "el destino más barato": se rechaza la hipótesis de que menor coste mensual implique mejor posición en el ranking (H1 ✗).

-Sí importa la capacidad económica local: a mayor poder adquisitivo, mejor valoración del destino (H2 ☒.

-La vivienda no resultó ser el factor decisivo frente a otras partidas: se rechaza que el coste de vivienda sea "lo más relevante" por encima de alimentación/cesta (H4 ✗).

-Mejor equilibrio coste/score detectado en destinos concretos: el análisis identifica países con buen balance "menos coste mensual + buen score" (p. ej., Rumanía, Hungría, Lituania, Tailandia, etc.).

-Seguridad vs. visado: la seguridad no aparece como clave para el ranking (H8 ✗), mientras que la facilidad para conseguir visa sí (H9 ☒.

10.2 Trabajo futuro (mejoras propuestas): -Actualización y robustez de datos: incorporar refresh periódico de fuentes y control de calidad (valores faltantes, outliers, criterios homogéneos) para mitigar desactualización.

-Mayor granularidad (ciudad/barrio): extender el análisis a nivel ciudad y, si fuese posible, a nivel intraurbano (barrios), para decisiones más accionables.

-Modelo de decisión multivariable: construir un índice ponderado configurable por perfil (prioriza coste, conectividad, seguridad, visado, bienestar) y comparar resultados por pesos.

-Segmentación de destinos: clustering para identificar "tipos" de destinos (barato+rápido internet, caro+alta calidad de vida, etc.).

-Producto final (visualización): dashboard con filtros por continente/métricas y comparador de destinos, basado en las variables integradas en el merge.

-En conclusión, el EDA indica que para destinos atractivos pesan más conectividad, bienestar y poder adquisitivo que 'ser el más barato'; y la vivienda no fue el factor decisivo frente a otros gastos (H4 ✗)" y proporciona una base cuantitativa para priorizar destinos según conectividad, bienestar y poder adquisitivo (y, como apoyo, estructura de gasto mensual y comidas fuera).

10.3 Consideraciones éticas y sostenibilidad del nomadismo digital (hallazgo adicional)

Uno de los aportes complementarios del estudio no se centra únicamente en la conveniencia para el nómada digital, sino en los posibles efectos socioeconómicos sobre las comunidades receptoras. Para ello, se propone

una métrica exploratoria de Responsabilidad Social, denominada Housing–Salary Ratio (HSR), definida como la relación entre el alquiler mensual estimado y el salario medio local.

Los resultados muestran la existencia de destinos en los que el alquiler mensual asociado a perfiles internacionales puede representar valores extremadamente altos respecto al salario local (por encima del 100% e incluso superiores en casos puntuales), lo que sugiere escenarios de alta presión de asequibilidad. En este tipo de contextos, la llegada de población con ingresos significativamente mayores puede intensificar dinámicas de segmentación de mercado y contribuir a procesos asociados a gentrificación o desplazamiento residencial.

Como implicación práctica, se recomienda que cualquier herramienta de recomendación de destinos (plataforma, informe o consultoría) incorpore esta señal como un indicador de riesgo social: (i) advertir al usuario cuando el HSR supere determinados umbrales, y/o (ii) desaconsejar activamente destinos con ratios especialmente elevados, priorizando criterios de sostenibilidad y equilibrio con la economía local.

Este enfoque abre líneas futuras de trabajo, como el diseño de un sistema de alertas y modelos predictivos (p. ej., machine learning) para anticipar presión de vivienda e identificar destinos con mejor equilibrio entre atractivo para el nómada y sostenibilidad social.