|  |  |
| --- | --- |
| ¿ES VIABLE HACERSE NÓMADA DIGITAL EN ESPAÑA? | Descripción breve  Un nómada digital es un profesional que utiliza las nuevas tecnologías para trabajar y lleva a cabo un estilo de vida nómada. Por lo general, trabajan de forma remota, en lugares como su hogar, cafeterías o bibliotecas públicas, en vez de hacerlo en un lugar de trabajo fijo. ¿Te atreves a unirte a la tribu global?  Torao Ogayar, Reiche Javier  DataScience |

Índice

1. Introducción y Contexto

2.Metodología

3. Análisis de Datos

4. Conclusiones

**1. Introducción y Contexto**

Objetivo del proyecto:

En la era actual, caracterizada por la creciente digitalización y flexibilidad laboral, el análisis de ofertas de trabajo en remoto adquiere una relevancia singular. Este proyecto se centra en examinar las oportunidades de empleo en remoto, un factor crítico y, a menudo, el primer paso indispensable para aquellos que aspiran a convertirse en nómadas digitales. La posibilidad de trabajar desde cualquier lugar no solo redefine las fronteras tradicionales del espacio de trabajo, sino que también es un prerrequisito fundamental para adoptar un estilo de vida nómada digital. Dado este contexto, nuestro análisis busca ofrecer una visión clara y detallada del panorama actual de las ofertas de empleo en remoto, proporcionando así una herramienta esencial para quienes buscan embarcarse en la aventura del nomadismo digital, una tendencia cada vez más prominente en el mercado laboral global.

La evolución del mercado laboral en los últimos años ha sido profundamente influenciada por un cambio significativo hacia el trabajo en remoto. Este cambio no es solo una respuesta a las circunstancias globales, como la pandemia, sino también un reflejo del creciente deseo de las personas por una mayor autonomía y bienestar en su vida laboral. La tradicional "carrera de ratas" y la monotonía de la oficina han dado paso a una preferencia por estilos de trabajo que favorecen la flexibilidad, el equilibrio entre la vida personal y profesional, y el bienestar mental. Este nuevo paradigma laboral, donde se valora más el control personal y la autodeterminación, ha llevado a un aumento en la demanda de oportunidades de trabajo en remoto a nivel global. La capacidad de trabajar desde cualquier lugar no solo ofrece una sensación de libertad y autonomía, sino que también se alinea con un reajuste de prioridades donde el bienestar personal se convierte en un factor clave. Esta tendencia emergente no solo redefine las expectativas y preferencias de los trabajadores, sino que también plantea nuevos retos y oportunidades para las organizaciones en todo el mundo.2.

**2.Metodología**

Recopilación de datos:

Los datos de los que nos nutrimos para el desarrollo de este trabajo proceden de ‘kaggle.com’ y del Instituto Nacional de Estadística (INE), de ahí hemos descargado los siguientes datasets:

* ‘alojamientos-turísticos -sample.csv’ alojado en la carpeta 3.Código\src\1.data \‘alojamientos-turísticos -sample’
* ‘datos\_agregados.csv’,’24900.’ y ’25198’.Alojado en la carpeta 3.Código\src\1.data\2.ciudades\_para\_teletrabajar(INE\_2022)
* ‘Dist\_entr\_ciudades.csv’,’jobs .csv’,’nomad\_cities.csv’,’pobla .csv’,’poblaciones .csv’,’trab\_remo.csv’

No todos los datos recopilados han sido empleados en el desarrollo de este trabajo ya que a medida que se ido implementando se han ido desechando aquello que no aportaba nada o no estaba alineado con los objetivos establecidos

Herramientas utilizadas:

En el desarrollo de este análisis de datos sobre nómadas digitales , se emplearon varias herramientas de software, cada una desempeñando un papel crucial en diferentes etapas del proceso:

Python: El lenguaje de programación principal utilizado, conocido por su versatilidad y amplio soporte para análisis de datos.

Pandas: Una biblioteca de Python fundamental para la manipulación y el análisis de datos. Se utilizó para la limpieza, transformación y exploración de los datos, permitiendo operaciones eficientes en los conjuntos de datos.

NumPy: Otra biblioteca de Python esencial, utilizada principalmente para cálculos numéricos. NumPy proporciona soporte para matrices y operaciones matemáticas de alto rendimiento.

Seaborn y Matplotlib (Pyplot): Herramientas de visualización en Python. Seaborn, que trabaja en conjunto con Matplotlib, fue usado para crear gráficos estadísticos informativos y atractivos, fundamentales para el análisis exploratorio de datos.

Beautiful Soup: Una biblioteca de Python empleada para el web scraping, que fue útil para recopilar datos de fuentes en línea.

Markdown: Utilizado para la documentación y presentación de análisis. Facilita la creación de informes legibles que combinan código, resultados y análisis.

CSV y Excel: Formatos de archivo para la importación y exportación de datos. Estos formatos son ampliamente utilizados para el almacenamiento y manejo de conjuntos de datos en forma tabular.

Archivos de Texto (TXT): Utilizados para el manejo de datos en un formato sencillo y accesible, especialmente útil para datos que no requieren estructuras complejas.

Cada una de estas herramientas contribuyó de manera significativa al proceso de análisis, desde la recopilación y preparación de los datos hasta su análisis y visualización final.

**3. Análisis de Datos**

Exploración inicial del conjunto de datos:

Lo primero que se llevo a cabo fue un análisis de los datos proporcionados Enel csv de trabajos en remoto (jobs.csv). Cuando se revisa la información vemos que los datos parecen parcialmente ‘cortados’ siguiendo el siguiente patrón : datos no muy vinculantes se informan al 100%, de importancia media al 50% y de importancia alta entre un 10-15%

Gráfico, Gráfico de barras

Descripción generada automáticamente

Esto no nos da mucho margen de movimiento por lo que me centro en alguno de los datos a ver si realizando una observación partiendo del principio de Pareto puedo hallar indicios de ese 20% de datos que me ayudan a encontrar un patrón que este detrás de este conjunto de datos.

El principio de Pareto, también conocido como la regla 80/20, describe un fenómeno que establece que aproximadamente el 80 % de los resultados provienen del 20 % de las acciones.

Ejemplo de Ley de Pareto

El 80% de los resultados de tu trabajo lo obtienes invirtiendo un 20% de tu tiempo. El 20% de la ropa de tu armario la usas el 80% de tu tiempo; y el 80% de la ropa restante solo la usas un 20% de veces

En línea con esta primera acción nos centramos en una de las variables que más se han informado para ver que información puedo extraer de ella . Esta variable es la variable ‘Provincia’ , ya que la de ciudad (city) es ambigua, al igual que otras como ‘sectors’, ‘expertise’,’type’ o ‘functions’.

Gráfico, Gráfico de barras

Descripción generada automáticamente

A continuación se revisa la estructura del data set para ver si puedo extraer mas información.

A pesar de disponer de pocos datos se opta por ver cual es la tendencia retributiva de l mercado y establecemos un análisis por escenarios:

Poco optimista y optimista, segmentando por alto-medio -bajo. Se considera poco optimista cuando entras a trabajar y te pagan lo mínimo , y optimista si te pagan lo máximo que se paga de inicio cuando te contratan. Así mismo se establecen los escenarios bajo-medio-alto según el Q25 , el Q50(mediana) y el Q75.

Para los resultados obtenido se realiza una corrección para pasar todos los salarios a una retribución dineraria de bruta y anual , ya que en las ofertas aparecen los salario por hora , por semana, por mes y por años.

Así mismo como al final nos interesa saber los que ganamos limpio al mes para saber cual es el presupuesto con el que suelen contar los nómadas digitales aplicamos tambien las retenciones por tramos del irpf que nos aplicaría la Agencia tributaria.

A continuación, se detallan los tramos del IRPF para 2020:

Con ingresos hasta 12.450 €, la retención será del 19 %

Desde 12.450 € hasta 20.200 €, se retendrá el 24 %

Entre 20.200 € y 35.200 €, será el 30 %

De 35.200 € a 60.000 €, el 37 %

Sobre los 60.000 €, la retención será del 45 %

Esto nos proporciona los siguientes resultados:

Imagen que contiene Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente con confianza media

A continuación se intenta ligar la información con otras fuentes relacionadas con la población en provincias y con la AEMET (se logra conectar con la API de AEMET, pero la información que ofrece requiere de mucho trabajo de la información, se valora y se abandona) . Estas acciones no fructifican y se pasa a analizar los datos de los dataset de los alojamientos turísticos alojamientos-turisticos-sample.csv para ver que poblaciones son las mas demandadas y si podemos hallan alguna relación con la del primer dataset.

De este dataset obtengo los precios que habitualmente se suele pagar por noche en las localizaciones mas demandadas, que al fin y al cabo será las que buscarán los nómadas digitales. Se hace desglose por meses y a nivel anual.

Tabla

Descripción generada automáticamente

Interfaz de usuario gráfica, Gráfico, Aplicación

Descripción generada automáticamente

Tabla

Descripción generada automáticamente

Finalmente se busca encontrar un patrón que nos diga si ese nómada digital se podrá permitir vivir sin problemas y saber que le puede esperar como norma general en las ciudades más interesantes de España . En primer lugar y basándonos en una información suministrada por el INE vemos que cuando de supera el umbral de 34% de nuestro presupuesto en vivienda comienza a ser un punto critico para cualquier español y eso incluye a los nómadas digitales. Cuando llevamos a cabo un análisis por ciudades , nos damos cuenta de que la matriz de correlación no sale bien cuando combinamos el dataset de alojamientos y el de nómadas digitales debido a que los valores de la segunda tabla son siempre constantes y eso genera un NaN en la matriz de correlación de Pearson . Este hándicap hace que me replantee la pregunta, no sobre qué me puedo esperar en una ciudad determinada sino que patron me puedo esperar en todas. Esto me da como resultado la siguiente matriz de correlación de cuyo estudio puedo sacar alguna conclusión y/o una serie de hipótesis para poder llevar a cabo análisis inferenciales que arrojen luz sobre algunas dudas vinculadas con nuestro estudio original y que sean de interés para nuestros objetivos.

Imagen que contiene Escala de tiempo

Descripción generada automáticamente

Limpieza y Preprocesamiento de Datos: Se ha tenido que manipular datos para eliminar valores NaN que nos aparecían en nuestro dataframe , así como aplicar el uso de máscaras para filtrar información.

Por último, tambien destacar la utilización fallida de webscraping con Beautiful Soup para descargar la grafica obtenida de los datos del INE que habla de los gastos del presupuesto de los españoles. Esta se encuentra en la pagina

url="https://www.eldebate.com/economia/20220630/asi-gastaron-dinero-espanoles-2021.html"

4. Conclusiones

El que sea viable o no nuestra aventura como nómada digital dependerá no ya de las probabilidades que se tienen a priori sino de la manera en que materialicen y que opciones salariales se tengan. Como resultado de esto nos podremos plantear si alcanzamos o no a pagar el gasto en vivienda , o si nos arriesgamos a superar ese 34% de gasto en vivienda que en principio no parece ser lo más sensato. También podremos decidir si nos compensa o no estar en España o irnos a otro sitio si no nos gusta el perfil de las ciudades españolas según los datos que nos aparecen en la matriz de correlación.