



Proyecto EDA

Cristina Cejas Sánchez



Chekiñ



Sobre Chekin



Chekin es una start-up que ha creado un software enfocado a hoteles y apartamentos turísticos.

A través de este software se registra la información de los huéspedes a la hora de realizar el check-in, agilizando su proceso en la recepción o incluso procesando el registro antes de la llegada al alojamiento.

Se basa en una suscripción mensual o anual que puede ir aceptando mejoras según las funcionalidades que desee el cliente.



La Hipótesis



¿EL USO DEL PRODUCTO Y QUIÉN LO UTILIZA TIENE
UN EFECTO EN EL CRECIMIENTO DE LA EMPRESA?

Subhipótesis 1

¿Ha aumentado el número de check-in online después del COVID?

Subhipótesis 2

¿Hay diferencias en el tipo de uso de Chekin según la características del huésped?

Subhipótesis 3

¿Qué porcentaje de clientes realiza algún tipo de mejora en la funcionalidad de Chekin?

Subhipótesis 4

¿Hay alguna correlación entre las mejoras en las funcionalidades del software y los beneficios?

Los Datos



1

Obtención de datos en bruto

El primer gran problema encontrado a la hora de analizar los datos fue la inmensa cantidad de datos en crudo de los que disponía la empresa. Estos reportes fueron descargados a través del backend, así como otros me fueron dados por parte de los desarrolladores de la empresa.

2

Tipo de datos

Recopilé 4 csv que correspondían a las características particulares de los usuarios, la de los huéspedes, información sobre las viviendas que realizaban el check-in y los beneficios de los últimos años, los cuales hubo que limpiar de manera concisa.

3

Limpieza de datos para poder hacer matching

Para poder relacionar cada csv, filtré por número de ID o e-mail. En el caso de los huéspedes me encontré con 5 millones de datos (que filtré por usuarios únicos), y por parte de los usuarios había registros de más de 16000 clientes (de los cuales estudié solo aquellos que a fecha actual estaban activos en la plataforma).

4

Matching

Finalmente fui mergeando los diferentes reportes para obtener datos específicos a estudiar según necesidad, añadiendo ratios para su mejor visualización posterior.

5

Visualización

Para mostrar los datos mediante gráficas utilicé las herramientas Tableau y Plotly

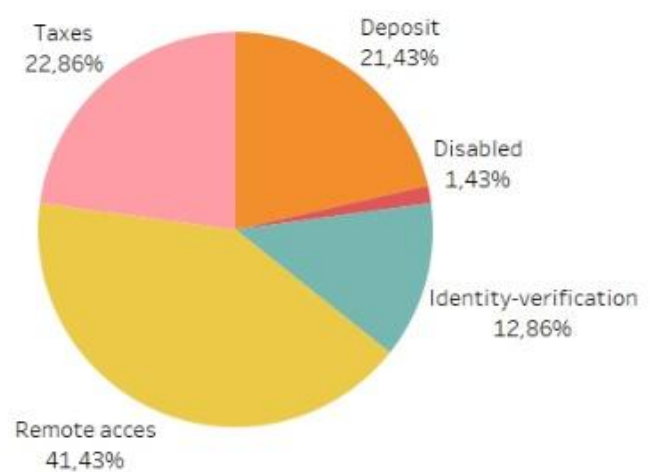
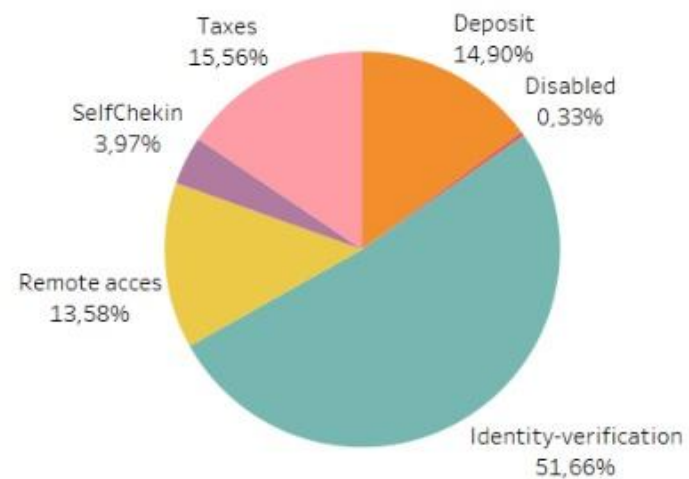
✓ 0.7s

| | | id | updated_at | status | type | current_period_items_data |
|-------|--------------------------------------|-------------------------------|------------|--------|--|---------------------------|
| 0 | c8099037-46ad-43af-a788-6b9558aaf801 | 2021-10-12 12:29:29.166 +0100 | CANCELED | HOU | [{"plan": "VR-Basic-monthly", "quantity": 1}] | |
| 1 | 313db5da-1c6b-4a6a-9800-295d980af81c | 2021-09-17 10:25:14.923 +0100 | CANCELED | HOU | [{"plan": "vacrent_basic_monthly", "quantity": "... | |
| 2 | 73c7e7aa-9269-493c-a7e0-d5c664ac439b | 2021-10-22 15:08:07.618 +0100 | CANCELED | HOT | [{"plan": "price_1HYxefBdXXC1yuNcpP8NSbIR", "q... | |
| 3 | d87270b5-a796-4ea9-b216-52d715668434 | 2022-10-11 12:22:01.723 +0100 | UNPAID | HOU | [{"plan": "vacrent_basic_monthly", "quantity": "... | |
| 4 | b455d53d-a64b-48c6-9254-5fa2e21df5b2 | 2021-09-17 10:26:45.348 +0100 | CANCELED | HOU | [{"plan": "vacrent_basic_monthly", "quantity": "... | |
| ... | ... | ... | ... | ... | ... | |
| 16095 | 5dd9ce77-60a7-4cbb-9b4a-f0f6ebe81e44 | 2022-10-11 12:49:10.547 +0100 | ACTIVE | HOU | [{"plan": "VR-Basic-monthly", "quantity": 30}] | |
| 16096 | 21691cd1-2c34-4c5d-802e-2af4a69384cb | 2022-10-12 01:58:14.852 +0100 | ACTIVE | HOU | [{"plan": "VR-Basic-free-monthly", "quantity": "... | |
| 16097 | 994a7b3e-79ea-4f51-bb60-04f768f96550 | 2022-10-13 21:37:39.361 +0100 | ACTIVE | HOT | [{"plan": "hotel_basic_monthly_2_1", "quantity": ... | |
| 16098 | 058dd6d0-649c-448d-a9ea-03053f4c9b9d | 2022-10-19 07:29:28.485 +0100 | ACTIVE | HOU | [{"plan": "VR-Basic-free-monthly", "quantity": "... | |
| 16099 | dcc9898b-06cd-4e65-8653-0d3abc03f199 | 2022-10-19 16:52:46.343 +0100 | ACTIVE | HOU | [{"plan": "VR-Basic-free-monthly", "quantity": "... | |

Ejemplo de un dataframe.
Muestra el tipo de suscripción a la que está sujeto el cliente.

| id | status | type | Email | number_of_properties | date | plan1 | quantity1 | plan2 | quantity2 | plan3 | quantity3 | plan4 | quantity4 |
|-------------------------------------|--------|------|----------------------------------|----------------------|------------|-----------------------|-----------|-----------------------|-----------|------------|-----------|------------|-----------|
| 90bc3adf-5f11-4039-a534-025e8ffb0cc | ACTIVE | HOT | e.knops@yays.com | 1 | 2022-10-05 | Basic | 26 | Identity-verification | 26 | Sin cambio | 0 | Sin cambio | 0 |
| ced5014c-6460-475e-8a1a-548ad5f91fe | ACTIVE | HOU | stay@biteintomadrid.com | 3 | 2022-08-31 | Basic | 3 | Sin cambio | 0 | Sin cambio | 0 | Sin cambio | 0 |
| f2199312-2c40-4e95-b87d-ce48cdd9a3 | ACTIVE | HOU | apartamentosusisevilla@gmail.com | 2 | 2022-03-31 | Identity-verification | 1 | Basic | 2 | Sin cambio | 0 | Sin cambio | 0 |
| 51799acd-6184-47ba-a4af-969e8aa32c | ACTIVE | HOU | elfaro1947@gmail.com | 4 | 2022-03-31 | Basic | 3 | Sin cambio | 0 | Sin cambio | 0 | Sin cambio | 0 |

En este caso vemos el dataframe una vez mergeado con otros dataframes.



Ejemplo de dibujo de cómo se distribuyen los planes de los suscriptores a las funcionalidades extras de Chekin

El

Análisis



1

Clientes utilizando checkin

Una vez obtenidos y limpiados los datos, se procedió a su estudio. Para ello empecé intentando entender quiénes eran los clientes de Chekin, qué tipo de funcionalidades eran las que más utilizaban y dónde tenían repartidas sus propiedades

2

Quiénes son los clientes de los clientes?

Por otro lado, empecé a evaluar a los huéspedes. De dónde venían, qué edad tenían y qué tipo de check-in realizaban, para así ver su evolución a lo largo de los años y si se podía encontrar alguna correlación entre esas características y el tipo de check-in realizado..

3

Propiedades

Posteriormente, filtré los usuarios por el número de propiedades que tenían, para así reconocer a los clientes más importantes de la empresa y estudiar las características que podían tener en común.

4

Beneficios

Finalmente analicé los beneficios de la empresa con el paso de los años para así extraer de dónde venían los mismos.

Las Conclusiones



La inmensa mayoría de los huéspedes provienen de España. No hay relación directa entre la edad o proveniencia de los huéspedes con el tipo de Check-in que realizan.



Desde la implantación del Check-in online en el 2019, claramente vemos como el crecimiento del uso del mismo ha sido exponencial hasta día de hoy (exceptuando el periodo de crisis dado por el COVID).



La empresa ha ido aumentando sus beneficios a lo largo del tiempo, encontrando su punto más alto actualmente, doblando los beneficios del año anterior en más de un 50%. De estos, más del 50% provienen de mejoras en algunas de las funcionalidades del check-in online, a pesar de que los upsellings solo representan el 20% de las suscripciones totales.



Por lo tanto, después de este análisis, podemos llegar a la conclusión de que aumentando el tipo de check-in a online, aumentamos la posibilidad de vender funcionalidades extras a los clientes, siendo estas las que más beneficios aportan a la empresa.

