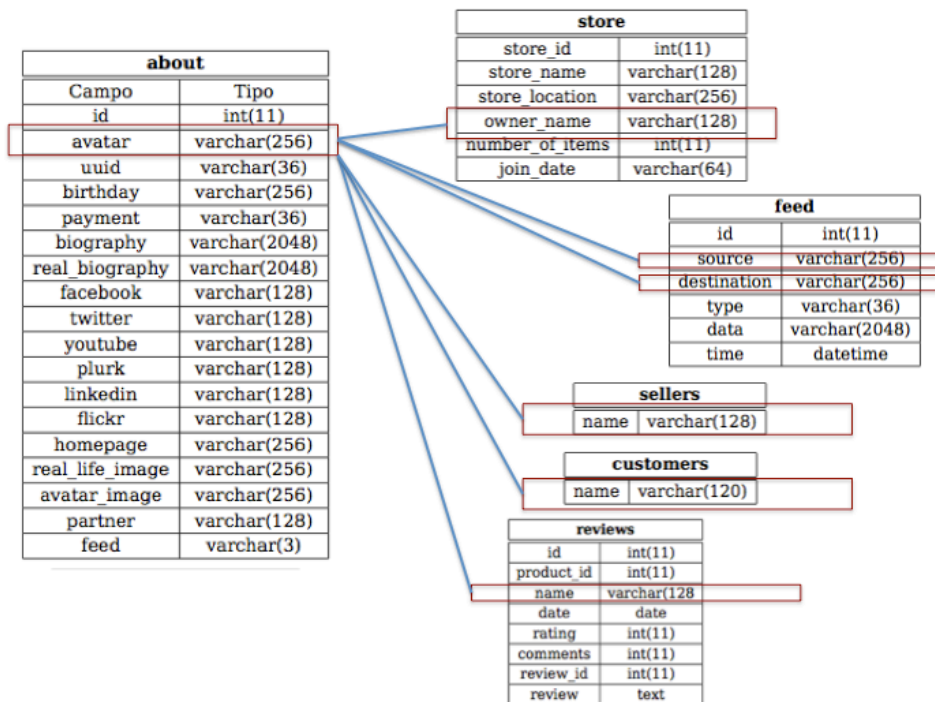


## Algunas indicaciones respecto a la tarea 2 de IIC 1005

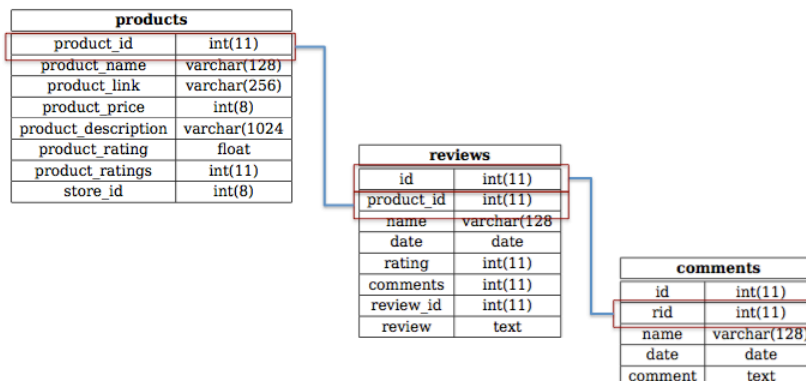
### 1. Base de datos

Para hacer joins entre las tablas, hay algunos campos que deben considerarse con cuidado:

Para hacer join de la table **about** (estos son los usuarios que tienen perfil en la red social mySecondLife) con tablas que tengan información de los usuarios en el marketplace (como **reviews**, **comments**, **customers** o **sellers**) se debe poner atención en los campos que hacen de identificador de usuario:



y para la relación entre **products**, **reviews** y **comments** (conectar reviews.id = comments.rid para obtener los comentarios sobre un review)



## 2. Minería de Datos

Respecto al uso de la regresión logística y su interpretación, los apuntes están disponibles en el siding (Carpeta Clases, archivo "[Week08\\_class15\\_ML.pptx](#)") pero aquí un resumen:

En la tarea,  $P$  = Probabilidad de que el usuario sea vendedor. Por otro lado, los coeficientes  $x_i$  representan variables (features) como el número de comentarios, la polaridad de los comentarios del usuario, etc.

$$\log\left(\frac{P}{1-P}\right) = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$$

En sk-learn, pueden obtener los coeficientes beta ( $\beta$ ) de un modelo de regresión usando el atributo `coef_` del modelo:

```
37 print logistic.get_params()
38 print logistic.coef_
39
```

Para interpretar los valores de los coeficientes beta, usar:

### Interpretar valores de $\beta$

- $\beta_j$  es la cantidad de cambio en logit por cada unidad que cambia  $X_j$
- A diferencia de regresión lineal, los  $\beta_j$  no se interpretan directamente, sino que  $\exp(\beta_j)$
- $\exp(\beta_j)$  : odds ratio (razón de disparidad)
  - $\exp(\beta_j) = 1$ , no hay cambio en los odds ratio
  - $\exp(\beta_j) < 1$  odds ratio decrece
  - $\exp(\beta_j) > 1$  odds ratio crece
  - $\exp(\beta_0)$ : baseline
- Ejemplo: en la tarea, si  $\exp(\beta_j) = 1.2$  y  $X_j$  es "número de amigos", por cada amigo adicional mi chance (odds) de ser dueño de tienda es 20% mayor.
- Recomendando revisar: [http://www.ats.ucla.edu/stat/mult\\_pkg/faq/general/odds\\_ratio.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/odds_ratio.htm)