# Capture Conversion Analysis

In credit risk modelling, **capture-conversion analysis** is essential for evaluating the effectiveness of delinquency triggers—such as 30+, 60+, or 90+ Days Past Due (DPD)—in identifying accounts that will eventually default over longer time horizons (e.g., within 24 months).

This dual analysis is critical for shaping **early warning systems**, optimizing **collections strategies**, and guiding **risk-based decisions**, such as account treatments or limit adjustments. By ensuring interventions are both **timely and targeted**, institutions can significantly reduce losses while enhancing operational effectiveness.

**Capture rate** measures the proportion of the eventual bad population that is flagged by each early delinquency definition

Capture Rate ​=

**For a given MOB value, as DPD increases**, the numerator gets smaller (fewer accounts reach target DPD), so **capture rate tends to decrease.**

**Conversion analysis** assesses how many of those flagged accounts actually transition into bad outcomes.

Conversion Rate ​=

**For a given MOB value, as DPD increases, the denominator shrinks faster than the numerator (since severe delinquency is rare but serious), so conversion rate tends to increase.**

Together, these metrics provide a holistic view of each trigger’s **predictive power and efficiency**, balancing early risk detection with precision.

Let eventual bads be defined as 180+ DPD in 18 MOB.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DPD** | **Accounts with DPD** | **Eventual Bads Captured** | **Capture Rate** | **Conversion Rate** |
| 30+ DPD 9MOB | 1,000 | 600 | 80% | 60% |
| 60+ DPD 9MOB | 600 | 500 | 67% | 83% |
| 90+ DPD 9MOB | 300 | 270 | 36% | 90% |
| **Total Eventual Bads** | — | 750 | — | — |

Based on the table above, short-term targets can be selected depending on how early delinquents need to be identified, balanced against the trade-off between capture rate and conversion rate.

# Code

def get\_cap\_conv\_rate(df,target\_cols,long\_target)

list1=target\_cols

list3 = []

list4 = []

list5 = []

final\_target = long\_target

for flag in list1

conversion\_rate = (df.filter((F.col(flag)==1) & (F.col(final\_target) ==1)).count())/(df.filter(F.col(flag)==1).count())

list3.append(conversion\_rate)

bad\_count = (df.filter(F.col(flag)==1).count())

total\_count = (df.count())

list4.append(bad\_count)

for flag in list1

capture\_rate = (df.filter((F.col(flag)==1) & (F.col(final\_target)==1)).count()) / (df.filter(F.col(final\_target)==1).count())

list5.append(capture\_rate)

data = list(zip(list1,list4,[total\_count]\*len(list1),list5,list3))

deptSchema = StructType([

StructField('flag',StringType(),True),

StructField('bad\_count',IntegerType(),True),

StructField('total\_count',IntegerType(),True),

StructField('Capture\_rate',DoubleType(),True),

StructField('Conversion\_rate',DoubleType(),True)])

df\_final = spark.createDataFrame(data, schema=deptSchema)

return df\_final

dpd\_months = [6,9,12,15,18,24]

exp2 = ['\*']\

+[f"case when max\_dpd{i}mob >= 30 then 1 else 0 end as targer\_{i}mob\_30\_flag" for i in dpd\_months]\

+[f"case when max\_dpd{i}mob >= 60 then 1 else 0 end as targer\_{i}mob\_60\_flag" for i in dpd\_months]\

+[f"case when max\_dpd{i}mob >= 90 then 1 else 0 end as targer\_{i}mob\_90\_flag" for i in dpd\_months]\

+[f"case when max\_dpd{i}mob >= 120 then 1 else 0 end as targer\_{i}mob\_120\_flag" for i in dpd\_months]\

+[f"case when max\_dpd{i}mob >= 150 then 1 else 0 end as targer\_{i}mob\_150\_flag" for i in dpd\_months]\

+[f"case when max\_dpd{i}mob >= 180 then 1 else 0 end as targer\_{i}mob\_180\_flag" for i in dpd\_months]

df = spark.sql("select \* from data.table")

df1 = df.selectExpr(exp2)

long\_target = "target\_24mob\_180\_flag"

target\_cols = [i for i in df1.columns if 'target' in 1]

df\_final = get\_cap\_conv\_rate(df1,target\_cols,long\_target)

# Appendix

**Days Past Due (DPD)** is a key metric in credit risk that indicates how many days a borrower has failed to make a scheduled payment after the due date. It’s a core component in measuring and managing delinquency and is often used to define risk thresholds:

* **30+ DPD**: The borrower has missed a payment by at least 30 days — often the first sign of financial distress.
* **60+ DPD**: Indicates more serious delinquency and increases the probability of default.
* **90+ DPD**: Commonly used as a proxy for default in many credit risk models, as the likelihood of recovery diminishes significantly.